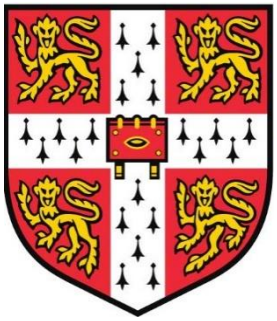


UNIVERSITY OF CAMBRIDGE

DEPARTMENT OF ENGINEERING

Density Functional Simulations of Defect Behavior in Oxides for Applications in MOSFET and Resistive Memory

Hongfei Li

St. John's College

This Thesis is submitted to the University of Cambridge for the Degree of Doctor of Philosophy

Declaration

This thesis contains the results of research performed by the author between October 2013 and December 2017 in the Department of Engineering, University of Cambridge. This dissertation is the result of my own work and includes nothing that is the outcome of work done in collaboration except where explicit acknowledgement is given.

This work has not been submitted in whole or in part for any other university degree or diploma.

Permission is granted to consult or copy the information and results contained within this thesis for the purpose of private study only, and not for publication.

This thesis contains about 35000 words.

Hongfei Li

Cambridge, December 2017

Dedicated to my parents

for their love, support

and encouragement

Acknowledgements

There are many people that I wish to acknowledge for their help during my PhD study in Cambridge. Without them, I would have no opportunity to finish my PhD study and this thesis. I would to express my biggest gratefulness to my supervisor, Prof. John Robertson. It is him who brought me to this exciting research field, led me through so many challenging research topics and provided the financial support for my conference trip and maintenance during my PhD. Prof. Robertson has always been patient, friendly, and helpful for discussion of my research progresses. He is also very concerned about my daily life and future, thus provided a lot of important advice and some of his own experience. His enthusiasm in research inspired me a lot to make every effort to making progress in my research life and career. I am very grateful for his support and help to publish papers and attending conferences. I benefit a lot from this process not only for scientific research, but also for personal development. As a supervisor, Prof. Robertson couldn't have done more to help me.

I would like to sincerely thank my colleagues in our research group, particularly to Dr. Yuzheng Guo, Dr. Huanglong Li, Dr. Xiaoming Yu, Dr. Xingyi Wu, Dr. Guofang Zhong, Dr. Zhigang Song, Mr. Haichang Lu, Ms. Shan Zheng, Mr. Guandong Bai, Mr. Yudan Zhao, Ms. Han Zhang for the discussion on my research project. Thanks for Prof. Nathan and Mr. Yang Su for the collaboration on the Cu_2O project. Thanks for Dr. Sedghi from the University of Liverpool for the collaboration on the ReRAM project. Thanks for Dr. Rui Huang from China for the collaboration on the SiCO project.

My Cambridge life could be dull without the accompany of so many cherished friends here: Xiaoming Yu, Bingyan Yang, Yuzheng Guo, Jieru Ying, Huanglong Li, Haichang Lu, Shan Zheng, Huihao Liu, Yunhui Zhang, Huiruo Li, Yanting Jin, Mao Mao, Zhi Jin, Dawei Sun, Jin Zhu, Yizhou Tan, Yi Man, Meng Zhang, Yu Song, An Bao, Thomas Choi, Xiaoxi Zhu, Yang Su, Sai Li, Ziqiang Feng, Xingkun Liang, Lu Cao, Yujiang Chen. You made my life in Cambridge full of joy and happiness.

Finally, I would like to thank my parents for their endless encouragement, help and love during my PhD study. I would like to also thank the Cambridge Trust, St. John's College, the Department of Engineering, Lundgren Fund for the financial support to cover my tuition, the maintenance, and the conference cost.

List of Publications

1. **H. Li**, Y. Guo and J. Robertson, Hydrogen and the Light-Induced Bias Instability Mechanism in Amorphous Oxide Semiconductors. *Scientific Reports*, **7**, 16858 (2017).
2. **H. Li**, Y. Guo, and J. Robertson, Face Dependence of Schottky Barriers Heights of Silicides and Germanides on Si and Ge. *Scientific Reports*, **7**, 16669 (2017).
3. **H. Li** and J. Robertson, Germanium oxidation occurs by diffusion of oxygen network interstitials. *Applied Physics Letters*, *110*(22), 222902 (2017).
4. **H. Li** and J. Robertson, Yttrium passivation of defects in GeO_2 and GeO_2/Ge interfaces. *Applied Physics Letters*, **110**, 032903 (2017).
5. **H. Li**, J. Robertson and Y. Okuno, Effect of metal oxide additions to quality on Ge/GeO_2 interfaces. *Journal of Applied Physics*, **120**, 134101 (2016).
6. **H. Li**, Y. Guo, J. Robertson, and Y. Okuno, Ab-initio simulations of higher Miller index Si: SiO_2 interfaces for fin field effect transistor and nanowire transistors. *Journal of Applied Physics*, **119**, 054103 (2016).
7. **H. Li**, Y. Guo and J. Robertson, AlN-GeO_2 based gate stack for improved reliability of Ge MOSFETs. *Microelectronic Engineering*, **147**, 168-170 (2015).
8. **H. Li**, Y. Guo and J. Robertson, Calculation of TiO_2 surface and subsurface oxygen vacancy by the screened exchange functional. *The Journal of Physical Chemistry C*, **119**, 18160-18166 (2015).
9. N. Sedghi, **H. Li**, I. F. Brunell, K. Dawson, Y. Guo, R. J. Potter, J. T. Gibbon, V. R. Dhanak, W. D. Zhang, S. Hall, J. Robertson and P. R. Chalker, Enhanced switching stability in Ta_2O_5 resistive RAM by fluorine doping. *Applied Physics Letters*, **111**, 092904 (2017).

10. Z. Lin[ξ], **H. Li**[ξ], R. Huang, Y. Zhang, J. Song, H. Li, Y. Guo, C. Song and J. Robertson, Defect Emission and Optical Gain in SiC_xO_y: H Films. *ACS applied materials & interfaces*, **9**, 22725-22731 (2017).
11. N. Sedghi, **H. Li**, I. F. Brunell, K. Dawson, R. J. Potter, Y. Guo, J. T. Gibbon, V. R. Dhanak, W. D. Zhang, J. Robertson, S. Hall and P. R. Chalker, The role of nitrogen doping in ALD Ta_2O_5 and its influence on multilevel cell switching in RRAM. *Applied Physics Letters*, **110**, 102902 (2017).
12. Y. Su[ξ], **H. Li**[ξ], H. Ma, J. Robertson and A. Nathan, Controlling Surface Termination and Facet Orientation in Cu_2O Nanoparticles for High Photocatalytic Activity: A Combined Experimental and Density Functional Theory Study. *ACS Applied Materials & Interfaces*, **9**, 8100-8106 (2017).
13. Y. Guo, **H. Li** and J. Robertson, AlN and Al oxy-nitride gate dielectrics for reliable gate stacks on Ge and InGaAs channels. *Journal of Applied Physics*, **119**, 204101 (2016).

ξ: The authors with this sign make equal contributions to this work.

List of Conference Presentations

1. **H. Li** and J. Robertson (Talk) Reduced Fermi Level Pinning with Silicides and Germanides, INFOS 2017, Germany, 2017
2. **H. Li** and J. Robertson (Poster) Ge Oxidation Does Not Follow the Deal-Grove Mechanism, INFOS 2017, Germany, 2017
3. **H. Li** and J. Robertson (Talk) Yttrium Passivation of Defects in GeO_2 and GeO_2//Ge Interfaces, MRS 2017 Spring Meeting, USA, 2017
4. **H. Li** and J. Robertson (Talk) Reduced Fermi Level Pinning with Silicides and Germanides, MRS 2017 Spring Meeting, USA, 2017
5. **H. Li** and J. Robertson (Talk) Yttrium Passivation of Defects in GeO_2 and GeO_2/Ge Interfaces, 47th SISC, USA, 2016
6. **H. Li** and J. Robertson (Talk) Ge Oxidation is not Deal-Grove, 47th SISC, USA, 2016
7. N. Sedghi, **H. Li**, I. Brunell, R. Potter, S. Hall, K. Dawson, P.R. Chalker, Y. Guo and J. Robertson, (Poster) Wider Memory Window in Ta_2O_5 RRAM by Doping, 47th SISC, USA, 2016
8. **H. Li**, Y. Guo and J. Robertson, (Talk) Rare Earth Element Doping of GeO2 for the Improved Interface Quality in Ge MOSFETs, WoDiM 2016, Italy, 2016
9. **H. Li**, Y. Guo and J. Robertson, (Talk) Rare Earth Element Doping of GeO2 for the Improved Interface Quality in Ge MOSFETs, MRS 2016 Spring Meeting, USA, 2016
10. **H. Li**, Y. Guo and J. Robertson, (Poster) Simulation of Higher Index Si:SiO_2 Interfaces for FinFET, MRS 2016 Spring Meeting, USA, 2016
11. **H. Li**, Y. Guo and J. Robertson, (Poster) Simulation of Higher Index Si:SiO_2 Interfaces for FinFET, 46th SISC, USA, 2015
12. **H. Li**, Y. Guo and J. Robertson, (Talk) AlN-GeO_2 Based Gate Stack for Improved Reliability of Ge MOSFETs, INFOS 2015, Italy, 2015

13. **Y. Guo**, H. Li and J. Robertson, (Talk) Comparison of Oxygen Vacancy Defects in Crystalline and Amorphous Ta_2O_5, INFOS 2015, Italy, 2015
14. **H. Li**, Y. Guo and J. Robertson, (Talk) Calculation of Interfacial Oxygen Vacancies in TiO_2 for Applications in CMOS, WoDiM 2014, Ireland, 2014

Contents

Abstract

Defects in the functional oxides play an important role in electronic devices like metal oxide semiconductor field effect transistors (MOSFETs) and resistive random-access memories (ReRAMs). The continuous scaling of CMOS has brought the Si MOSFET to its physical technology limit and the replacement of Si channel with Ge channel is required. However, the performance of Ge MOSFETs suffers from Ge/oxide interface quality and reliability problems, which originates from the charge traps and defect states in the oxide or at the Ge/oxide interface. The sub-oxide layers composed of Ge^{II} states at the Ge/GeO_2 interface seems unavoidable with normal passivation methods like hydrogen treatment, which has poor electrical properties and is related to the reliability problem. On the other hand, ReRAM works by formation and rupture of O vacancy conducting filaments, while how this process happens in atomic scale remains unclear. In this thesis, density functional theory is applied to investigate the defect behaviours in oxides to address existing issues in these electronic devices.

In chapter 3, the amorphous atomic structure of doped GeO_2 and Ge/GeO_2 interface networks are investigated to explain the improved MOSFET reliability observed in experiments. The reliability improvement has been attributed to the passivation of valence alternation pair (VAP) type O deficiency defects by doped rare earth metals. In chapter 4, the oxidation mechanism of GeO_2 is investigated by transition state simulation of the intrinsic defect diffusion in the network. It is proposed that GeO_2 is oxidized from the Ge substrate through lattice O interstitial diffusion, which is different from SiO_2 which is oxidized by O_2 molecule diffusion. This new mechanism fully explains the strange isotope tracer experimental results in the literature. In chapter 5, the Fermi level pinning effect is explored for metal semiconductor electrical contacts in Ge MOSFETs. It is found that germanides show much weaker Fermi level pinning than normal metal on top of Ge, which is well explained by the interfacial dangling bond states. These results are important to tune Schottky barrier heights (SBHs) for n-type contacts on Ge for use on Ge high mobility substrates in future CMOS devices. In chapter 6, we investigate the surface and subsurface O vacancy defects in three kinds of stable TiO_2 surfaces. The low formation energy under O poor conditions and the +2 charge state being the most stable O vacancy are beneficial to the formation and rupture of conducting filament in ReRAM, which makes TiO_2 a good candidate for ReRAM materials. In chapter 7, we investigate hydrogen behaviour in amorphous ZnO. It is found that hydrogen exists as hydrogen pairs trapped at oxygen vacancies and forms Zn-H bonds. This is different from that in c-ZnO, where H acts as

shallow donors. The O vacancy/2H complex defect has got defect states in the lower gap region, which is proposed to be the origin of the negative bias light induced stress instability.

Chapter 1. Introduction

1. MOSFET and Scaling

The transistor technology has been the heart of electronic industry for more than 50 years, which has also been regarded as one of the most important invention of the 20th century. In 1959, the pioneering work of Dawon Kahng et al from the Bell Labs gives birth to the invention of the metal oxide semiconductor field effect transistor (MOSFET) [1]. MOSFET was soon proved to be an excellent elenctronic switch, and got widely integrated into logic circuits.

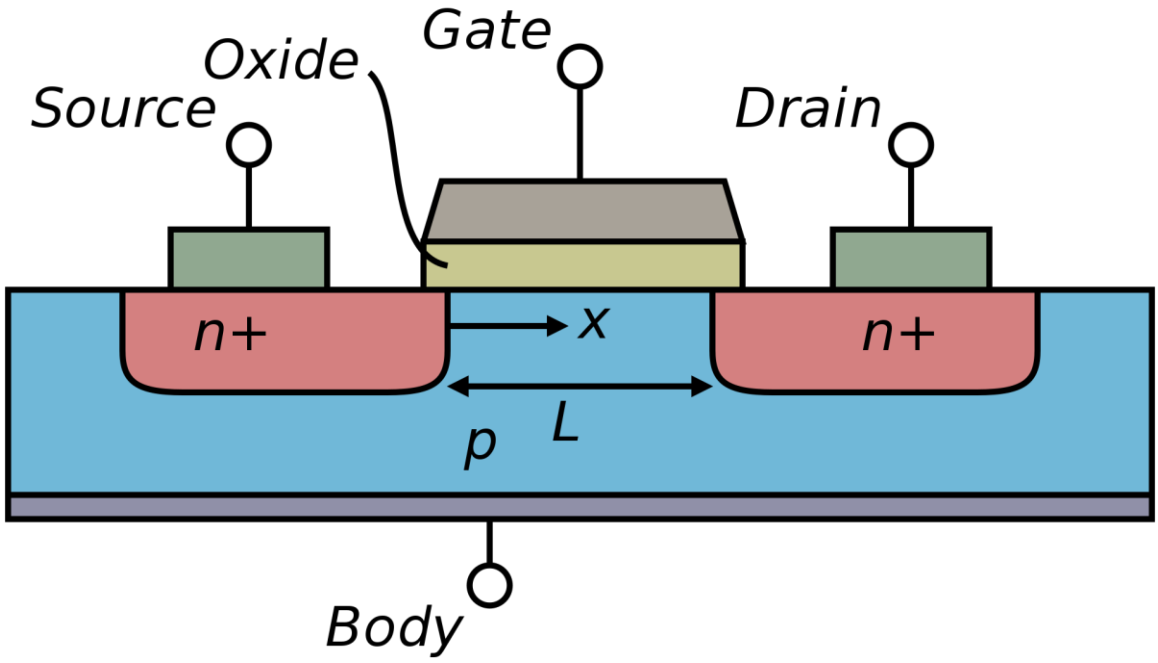


Fig 1.1 The device structure of a typical MOSFET device

The typical structure of the MOSFET is shown in **Fig 1.1** The semiconductor part is usually fabricated with Si, Ge, SiGe, or III-V materials, which acts both as the channel for charge carriers and also as the substrate. The Gate controls the on/off state of the device and is separated with the semiconductor channel by the insulator layer. The source and drain are connected by the channel, and usually made of highly doped semiconductor. The channel length and the oxide layer thickness are the two most important feature size for the MOSFET technology.

The continuous pursuit of more advanced integrated circuit leads to fast development in MOSFET technology. The well-known Moore's Law was put forward in 1965 which predicted the number of transistors doubles approximately every year. This empirical prediction was

further modified from each year to every 18 months, and included other criteria like doubling the chip performance, doubling transistor density, which successfully predicted the development trend of MOSFET for several decades.

The rapid development of the microelectronic industry has been driven by the idea of shrinking MOSFET device size linearly while keeping the device structure design unchanged, which known as the scaling of MOSFET. The initial scaling scheme was the "constant voltage scaling", which had the device size decrease in ratio and kept the device working voltage unchanged to make it compatible with early integrated circuit chips. It was however soon found to introduced increasing electric field within the device and resulted in breakdown. To deal with this problem, Dennard et al put forward the new scaling scheme at the International Electron Device Meeting (IEDM) 1972[2], which was further developed into the 'constant field scaling' scheme. The core idea was to decrease device working voltage and increase doping concentration when shrinking the MOSFET feature size, so as to keep the electric field in the device as constant. This new scaling scheme also led to the decrease of single device power and constant power density.

Table 1.1 Comparison of the constant voltage scaling and the constant field scaling.

Parameter	Constant Voltage Scaling	Constant Field Scaling
Gate Length	1/k	1/k
Gate Width	1/k	1/k
Electric Field	k	1
Oxide Thickness	1/k	1/k
Substrate Doping	k^2	k^2
Gate Capacitance	1/k	1/k
Oxide Capacitance	k	K
Voltage	1	1/k
Current	k	1/k
Power	k	$1/k^2$

The constant field scaling has been successful for the Si MOSFET until the sub 90nm technology [3]. The physical thickness of the SiO_2 layers has become too thin that the gate leakage current caused by electron quantum tunnelling can no longer be neglected. The tunnelling current through oxide layers increase exponentially when the oxide thickness is

decreased. This has prevented the further scaling as the power consumption due to the gate leakage current is too high.

Further scaling thus requires decreasing the device dimensions while not reducing oxide thickness. The compromise method is to keep oxide capacitance follow the scaling trend while not reducing oxide thickness. As the oxide capacitance is given by the following equation, the straight forward idea is to increase the dielectric constant of the oxide, which calls for the replacement of SiO_2 with high-K oxide.

$$C=K_0K_rA/t \tag{1.1}$$

Where K_0 is the absolute dielectric constant of the vacuum, K_r is the dielectric constant of the oxide, A is the surface area of the oxide, and t is the oxide's physical thickness.

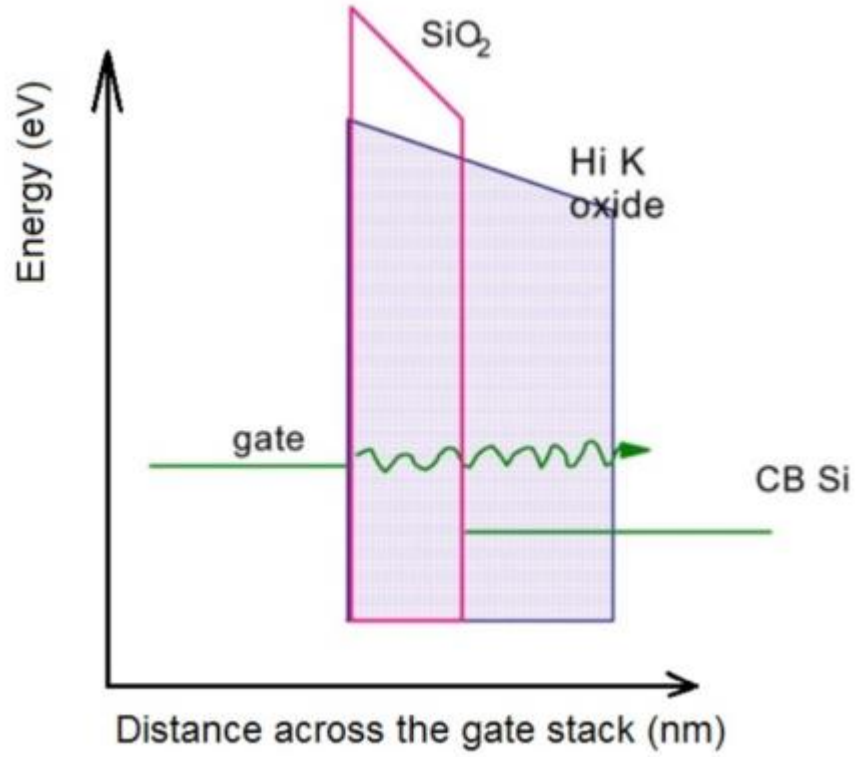


Fig 1.2 Illustration of how high-K gate oxide reduces the leakage current. Replace SiO2 with high K oxide enables to use thicker oxide, which effectively reduce the quantum tunnelling current. Figure reproduced from Ref[4]

The use of candidate oxide with higher dielectric constant than SiO_2 in MOSFET can effectively enable us to have oxide layer thick enough to supress tunnelling current while keeping oxide layer capacitance still follow the trend by constant field scaling scheme, as

shown in **Fig** 1.2. The other requirements for the oxide candidates include good interface quality with Si, and wide band gap as well as appropriate band alignment with Si for large enough band offset. However, oxide insulators with large dielectric constant usually have small band gap [4], as shown in **Fig 1.3**. The trade-off between dielectric constant and band gap has narrowed down the candidates such as HfO_2, Al_2O_3, $LaAlO_3$. The HfO_2 has been proved to quite suitable to Si technology and widely adapted.

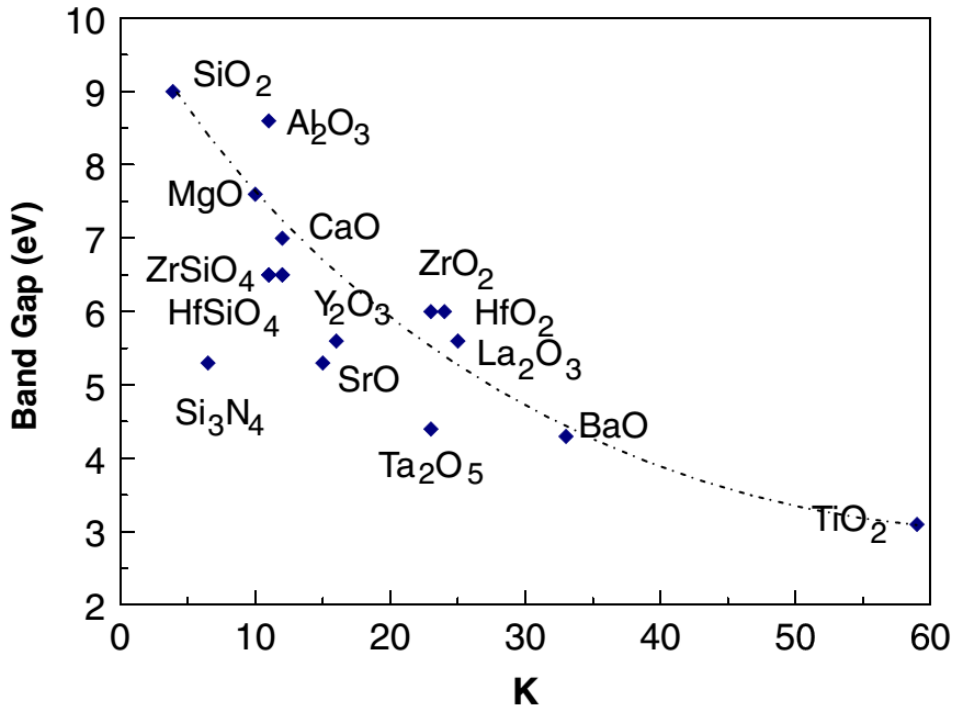


Fig 1.3 Illustration of trend of band gap and dielectric constant of oxides. [4]

2. Carriers Mobility for Channel Materials

With the further scaling of MOSFET as well as the substitution of SiO_2 with high-K oxide, there has been an intrinsic desire to replace the Si part with some high mobility channel materials. On one hand, the reason for the dominant position of Si technology in MOSFET is the quite good quality of the Si:SiO_2 interface. The SiO_2 layers exist in amorphous phase on top of the Si substrate, which well passivate all the dangling bonds at the Si surface. It is not difficult to obtain the Si:SiO_2 interface with quite low interfacial defect density on all of the traditional Si surface directions (001, 110, 111) [5-9]. As SiO_2 has been replaced with high-K oxide, Si is no longer the best option as the channel material. On the other hand, as the MOSFET feature size has been scaled down, problems including short channel effect, gate control, subthreshold conduction appear. It is natural to substitute Si channel with other high

mobility channel to continue the advance of device switching speed. The high mobility channel materials can have high carrier moving speed at same external electric field, which enables the MOSFET device to react faster at same applied voltage V_{dd}.

The mobility of semiconductor is an intrinsic parameter to the material and can be affected by many factors. When an external electric field is applied to a semiconductor, the electrons and holes will drift with certain velocity v. In addition, there is a linear relationship between the external electric field E and the carrier velocity v when the external field is not too large. The coefficient for such linear relationship is defined as the charge carrier mobility for this material, as shown in the equation below, where μ is the mobility.

$$v=\mu E \tag{1.2}$$

The mobility of semiconductor can be determined by several factors. First, how difficult it is to increase the carrier momentum p in semiconductor with applied external electric field, and this is usually described the effective mass of the carrier. This can be an analogous to the inertial mass of particle which determine the difficulty to accelerate the particle with external force.

From the energy band theory, the energy-wavevector relationship for electrons or holes follow the equation:

$$E = \hbar^2k^2/2m^* \tag{1.3}$$

It turns out to be that the inverse of the effective mass $1/m^*$ determine the curvature of the energy band [10]. For electrons near conduction band minimum and holes near valence band maximum region, the curvature of the energy band can be quite large, which suggest rather small effective mass to the order of 0.1 m_e (electron mass).

The effective mass for carrier can be different for different bands, see **Fig 1.4** for a prototype for different band types. Electrons in conduction band behave differently in direct and indirect band gap materials. For direct band gap materials, the conduction band is mainly composed of electron S states. On the other hands, for indirect band gap materials like Si and Ge, the conduction band can be either P type or a mixture of S and P type, which correlates to the transverse effective mass and longitudinal effective mass respectively.

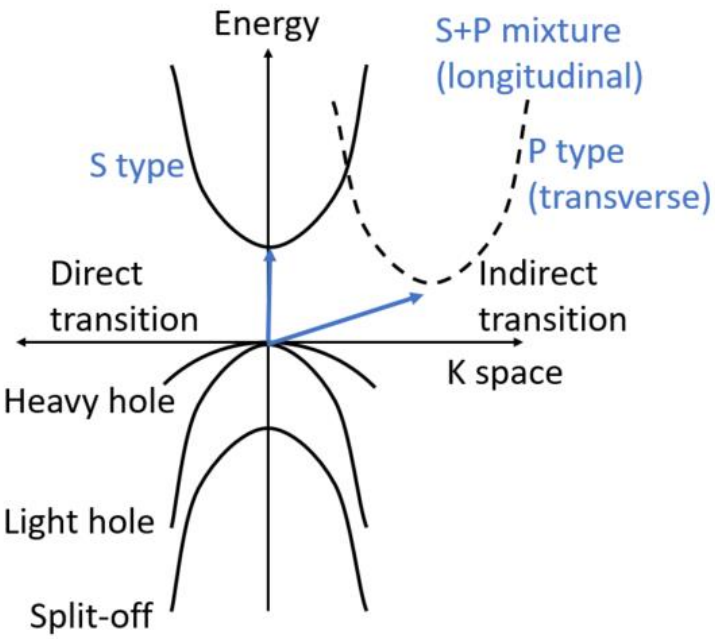


Fig 1.4 Valence and conduction band type in typical semiconductor

For holes in valence bands, there are usually more than one valence bands which are contributed by a few degenerate states. The curvature of these valences can differ greatly thus suggest different value of effective mass for different bands. The density of states related effective mass is calculated as the geometric mean of different type effective mass. The Drude model proposed the effective mass related to conductivity should be the harmonic mean of different effective mass, as the total energy is the sum of energy of carrier from each bands on each direction. The relationship between energy E, wavevector k, effective mass m^*, is given in the equation below (subscript i stands for different directions, which is important for anisotropic energy band minimum/maximum region like Si):

$$E = \hbar^2k^2/2m^* = \sum_{i=x,y,z} \hbar^2 k_i^2/2m_i^* \quad (1.4a)$$

$$m^* = \hbar^2/(\vartheta^2E/\vartheta k^2) \quad (1.4b)$$

The second factor which determines the mobility of semiconductor is the scattering effect. The carrier with high moments can be scattered by ionized impurity, acoustic phonon, optical phonon, defect scattering. The existence of any type of scattering can reduce the carrier mobility. The ionized impurities are usually activated dopants in semiconductor, which can react with charge carriers through Columbic reaction. Therefore, highly doped semiconductor can usually have lower mobility. The acoustic phonon scattering originates from the reaction between carriers and lattice vibration, which depends heavily on temperature. Optical phonon

exists at high external electric field, while defect scattering makes non-neglectable contribution when high density of defects appears in either bulk or surface, interface region.

Table 1.2 Effective mass and mobility of Si and Ge.

Property	Si	Ge
Indirect Band Gap (eV)	1.12	0.67
Direct Band Gap (eV)	3.2	0.80
Electron Effective Mass (m_e)	$m^*_l = 0.98$ $m^*_t = 0.19$	$m^*_l = 1.64$ $m^*_t = 0.082$
Hole Effective Mass (m_e)	$m^*_{LH} = 0.16$ $m^*_{HH} = 0.49$	$m^*_{LH} = 0.044$ $m^*_{HH} = 0.28$
Electron Mobility ($cm^2V^{-1}s^{-1}$)	1500	3900
Hole Mobility ($cm^2V^{-1}s^{-1}$)	450	1900

Table 1.2 shows the effective mass and mobility for electrons and holes in Si and Ge. The light hole and heavy hole effective mass are $0.16m_e$ and $0.49m_e$ in Si, while in Ge the numbers become $0.044m_e$ and $0.28m_e$ which are both smaller than in Si. This partly explains the higher mobility of 1900 $cm^2V^{-1}s^{-1}$ in Ge compared to 450 $cm^2V^{-1}s^{-1}$ in Si. Noted that Ge also has larger electron mobility than Si owing to the smaller transverse effective mass.

3. Ge MOSFET

Ge was actually the semiconductor used to fabricate the first MOSFET in the Bell Lab. Ge has been chosen for its good crystal quality in that early period [11]. However, the integrated circuit industry was soon dominated by Si technology owing to the good interface quality of $Si:SiO_2$ interface, which was suitable for fabricating MOSFETs. Recently, Ge has attracted the interest of MOSFET researchers again as a promising high mobility channel material.

As shown in **Fig 1.5**, Ge has the largest hole mobility among the common channel materials. Thus, much efforts and focus has been put on Ge as the substitution materials for Si channel. In addition, Si and Ge resides in the same chemical group so that they share many physical and chemical properties. Si and Ge has quite similar lattice structure and lattice parameter and their crystal structure and their native oxide. Despite the band gaps of the two semiconductors differ

by around 0.5 eV, the electron affinity of them are almost the same. Therefore, the criteria for choosing the high-K oxide for Ge are quite similar to that of Si. The feasibility of Ge processing in an existing standard Si product line has been explored and proved to be promising.

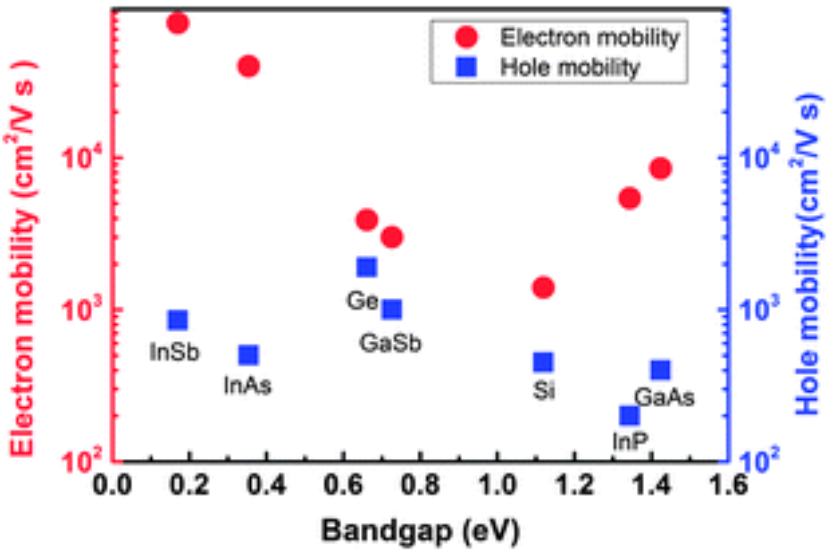


Fig 1.5 Electron and hole mobility of channel materials. Figure reproduced from Ref[12]

4. Oxide Selection for Ge MOSFET

Ge as a semiconductor is superior to Si for the higher carrier mobility, good crystal quality, and lower processing temperature. However, the choice of oxide compatible with Ge has been an unsolved problem. Unlike Si has good native oxide of SiO_2, Ge usually form poor interface with its native oxide GeO_2. The Ge:GeO_2 based stack has been reported to suffer from high interfacial trap density (D_{it}), poor reliability, poor resistance to high temperature and water [13-15].

The poor quality of of Ge:GeO_2 interface has been attributed to its suboxide GeO or GeO_x. It is known that Ge:GeO_2 interface can be made smooth and abrupt by careful preparation [16,17]. This interface can show low D_{it} in some cases as well [18]. However, the reliability of Ge stack is not as expected. The most important reaction lead to trouble in fabrication of Ge MOSFET is as follows:

$$Ge + GeO_2 \rightarrow 2GeO \quad \text{at} \sim 400\ ^\circ C \tag{1.5}$$

Such reaction can happen at the interface, consumes Ge from substrate and converts GeO_2 to sub-oxide. On the other hand, the Ge:GeO_2 interface is a pool for O vacancy. O vacancy can

diffuse from the interface to the GeO_2 bulk layers and make GeO_2 slowly evolves to GeO [19]. The reaction temperature is lower than the GeO_2 oxidation temperature and the dopants activation temperature, thus the formation of this sub-oxide seems difficult to be avoided [20]. Moreover, GeO_x is of poor electrical quality and volatile, thus detrimental to the Ge stack. Therefore, the slow evolution from GeO_2 to GeO together with the poor quality of GeO makes the Ge stack of poor reliability.

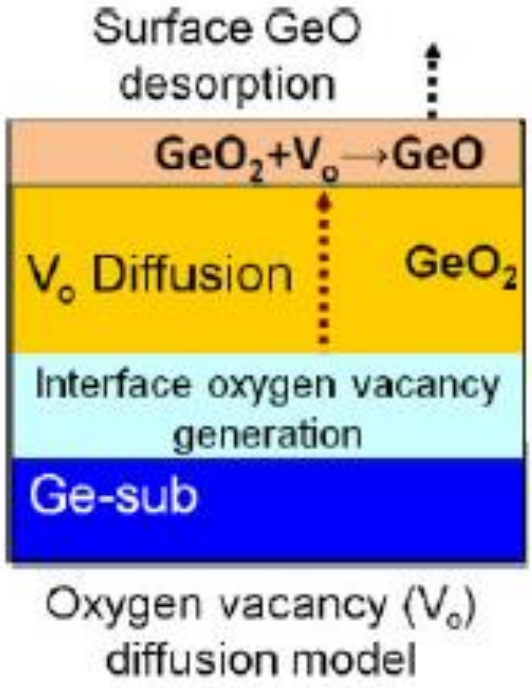


Fig 1.6 The oxygen vacancy diffusion model show GeO_2 evolves into GeO with the diffused oxygen vacancy from interface region. Figure reproduced from Ref[25]

The origin of GeO_2's poor reliability has been widely investigated. High density of P_b centres is found in Ge:GeO_2 stack through electron spin resonance measurement [21], see **Fig** 1.8. P_b centres can be Ge dangling bonds, which are thought to be related to poor reliability. The XPS result suggests multiple sub-oxide state exist in layers at the interface [22], as shown in **Fig** 1.7. Some literature reports Ge^{II} valence state is more stable in GeO_x than the equivalent Si^{II} in its oxide [23,24]. It is also reported that isotope tracer studies show the lattice O diffusion and the through the GeO_2 layer during the oxidation process, which is different from the situation in SiO_2 [25] (**Fig** 1.9). Some studies also suggest the Ge-O bond is less stable than Si-O thus GeO_2 amorphous can have less network rigidity compared to SiO_2 [26]. In addition, it is also noted that there exists a strange relationship between GeO_2 oxide thickness and the oxidation time (**Fig** 1.10) compared to the constant oxidation speed in SiO_2[27]. All these literatures

suggest that the oxidation mechanism in GeO_2 to be different from that in SiO_2, while the actual the mechanism remain to be unknown.

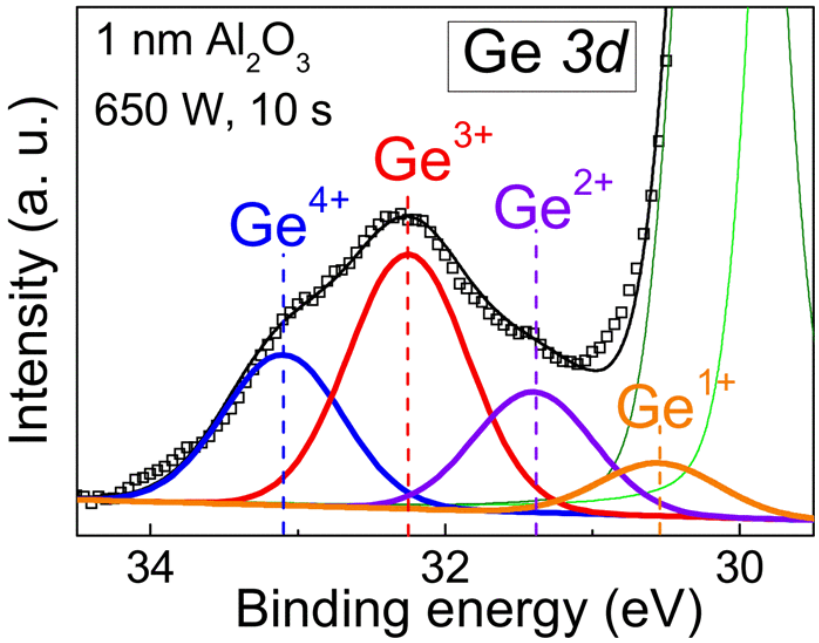


Fig 1.7 Deconvolution of the XPS spectrum of an Al_2O_3 GeOx/Ge structure representing the direct evidence of Ge sub-oxides existence. Figure reproduced from Ref[22]

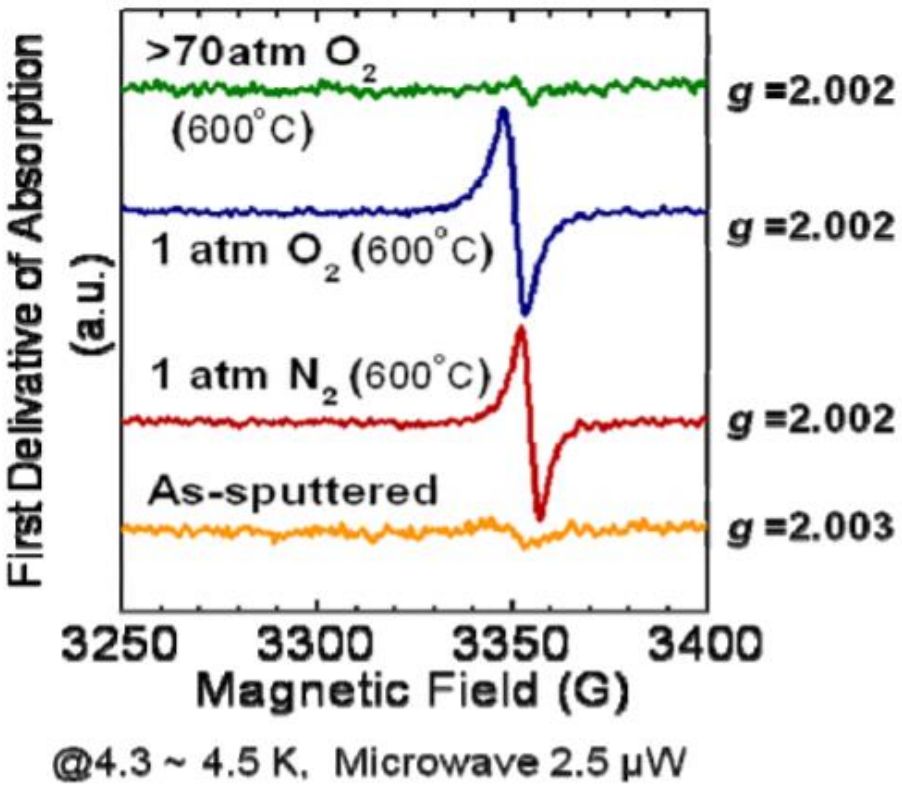


Fig 1.8 ESR signal detection from various GeO_2 films on Ge. Figure reproduced from Ref[21]

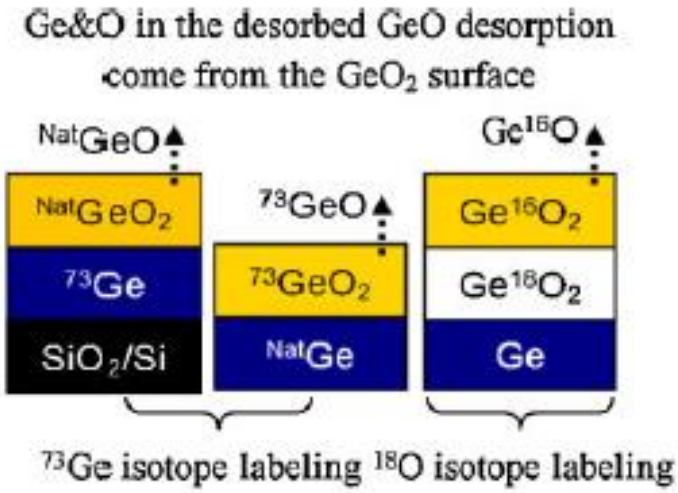


Fig 1.9 Schematic figure show isotope tracer study which confirm diffusion of O across GeO_2 layers. Figure reproduced from [25]

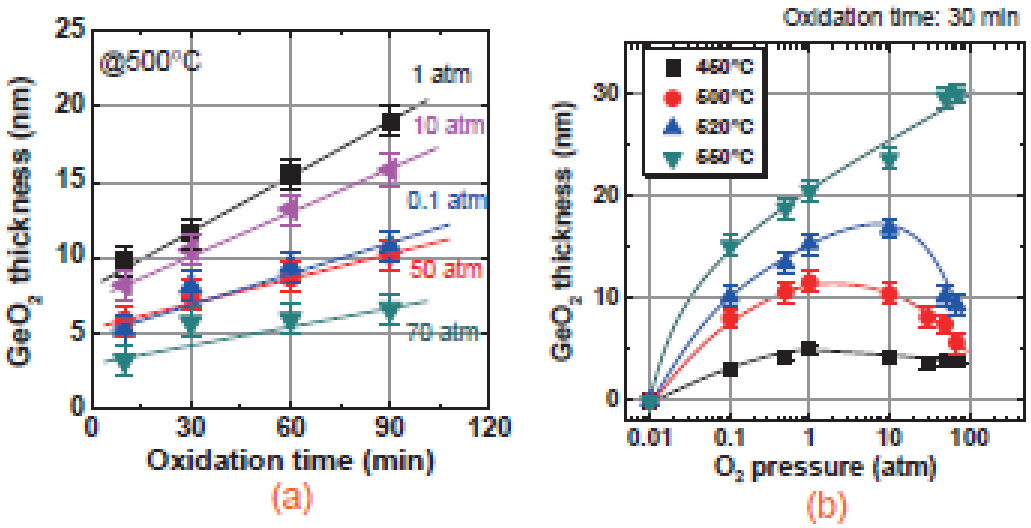


Fig 1.10. GeO_2 oxidation thickness with different oxidation time and oxygen pressure. Figure reproduced from Ref[27]

Another question is why GeO_2 usually cannot passivate Ge substrate well. Ge and Si are in same chemical group, and has quite similar crystal structure for their native oxide. The semicondoctor:native oxide stack for Si and Ge are both prepared by high temperature oxidation of Si(or Ge) substrate. However, the Ge:GeO_2 stack mostly has poorer interfacial quality and higher defect density in oxide layers. In addition, GeO_2 show strange dependence of oxidation rate [27]. As shown in the **Fig 1.10**, at high oxidation temperature of 550 °C,

higher oxygen pressure makes thicker GeO_2 layers on Ge substrate. On the other hand, low temperature oxidation shows the GeO_2 growth thickness increase first than decrease when vary the O_2 pressure across different sample. These findings raise the question that whether the oxidation mechanism for these two materials are same or not.

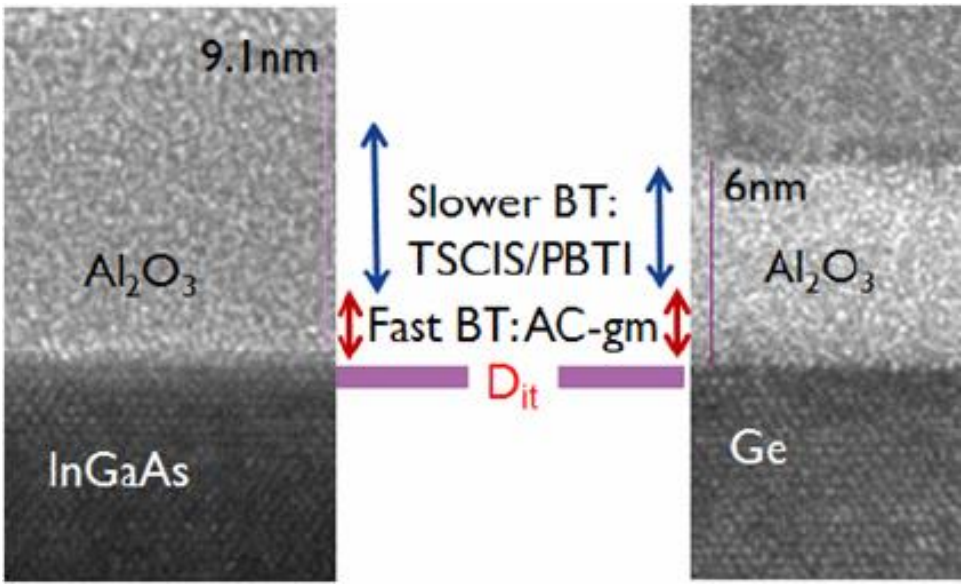


Fig 1.11 Cross sectional TEM images of the oxide-InGaAs and oxide-Ge interfaces to illustrate the distribution of border traps. (BT = border traps) Figure reproduced from Ref[28]

Other choice of oxide for Ge stack has been explored as well. HfO_2 works well on SiO_2 as a high-K oxide. However, HfO_2 react with Ge or GeO_2, and many built Ge:HfO_2 stack was found to form poor quality of $GeHfO_x$ layer due to Ge and Hf can diffuse into lattice structure of each other. On the other hand, Al_2O_3 has also been investigated as the potential oxide to Ge MOSFET as an oxygen diffusion barrier to prevent degradation of Ge:Oxide interface. It is however found that the Ge:Al_2O_3 gate stack suffer from a charge trapping at existing defects in their gate dielectrics, which shift the gate threshold voltage. Such degradation of has been shown to exist in III-V: Al_2O_3 stack as well. Due to the presence of Al_2O_3 in both cases, it is concluded that such bias instability originates from Al_2O_3. The instability comes from border traps across Al_2O_3 layers, which can also introduce high D_{it} [28]. Therefore, modification and improvement on gate dielectric of Ge:Al_2O_3 stack must be made to achieve the 10 year reliability life time.

5. Schottky Barrier and Fermi Level Pinning in Ge MOSFET

The performance of Ge MOSFET can also be limited by the contact resistances, which is due to the Schottky barriers at the contacts. A standard way to minimize the Schottky barrier heights(SBH) is to vary the contact metal, as the simple Schottky barrier model suggest the SBH is determined by semiconductor electron affinity and metal work function. This is however quite difficult in reality as the Ge SBH doesn't vary too much for metals with a large range of work function. As shown in **Fig 1.12**, for most metals the n type SBH is very large.

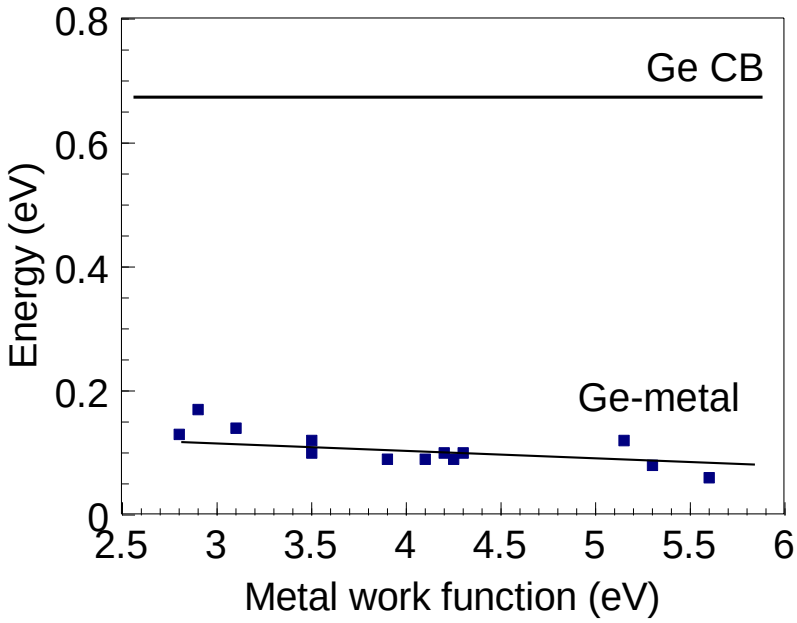


Fig 1.12 Schottky barrier height of Ge:Metal interface for metals with a large range of work function.

This is due to the well-known "Fermi level pinning" (FLP) effect. Heine developed the Metal Induced Gap States model to explain the FLP effect, which says the tails of metal wavefunction can expand into surface region of semiconductor and acts like surface states [29]. Tejedor et al further developed the model by introducing the concept of charge neutrality level (CNL): "the CNL is defined as an energy level below which interface charges at the semiconductor band gap are cancelled out by the lack of states in the semiconductor valence band" [30,31]. To be specific, the CNL can be influenced by surface states due to surface dangling bonds, and the band tails which expand into band gap region. The both kind of band gap states jointly determine the energy level for Fermi level so that the semiconductor surface can be locally charge neutral. When the Fermi level of the metal is not at the CNL of the semiconductor, the semiconductor side get locally charged and introduce an interfacial dipole. The dipole drives

the energy band realignment between metal and semiconductor, thus pins the Fermi level near CNL.

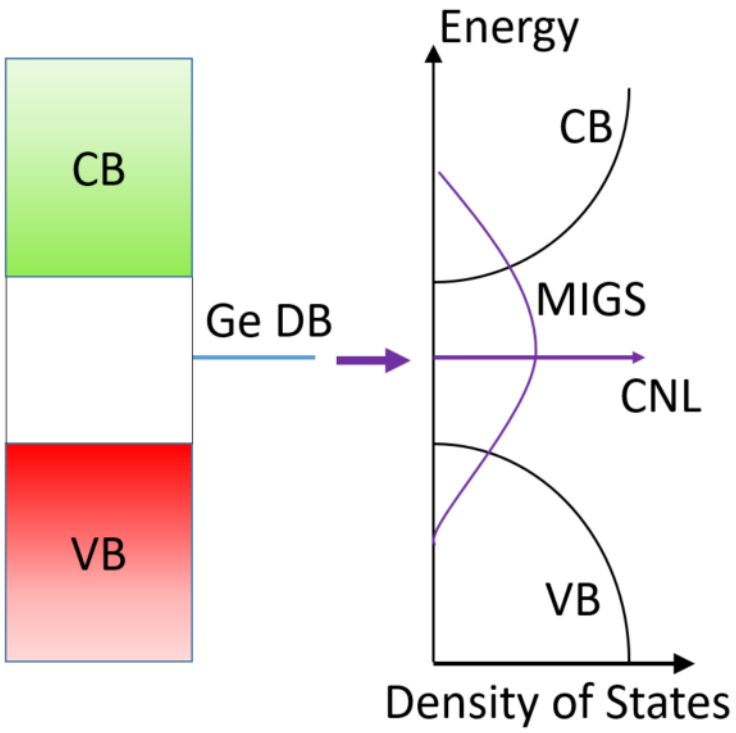


Fig 1.13 The schematic figure to show metal Fermi levels get pinned at the CNL.

6. Resistive memory

Rapid development of information technology calls for high-speed, low power, and nanoscale non-volatile memories. The conventional memory devices, including Flash memory, are faced with physical and technical limits during the scaling process [32, 33]. These conventional memories suffer from performance degradation such as leakage current and decreasing endurance when their scale is reduced to a certain range. The Resistive Random Access Memory (ReRAM) is believed to be able to circumvent the limitations of conventional memory devices [34] since its operation mainly based on 'resistive' change [33] rather than 'capacitive' change. Therefore, the ReRAM based on resistance switching (RS) phenomenon attracts much attention in recent years [35-38]. The key idea of the ReRAM relies on the dependence of electrical resistance of certain materials (mainly transition metal oxide and Perovskite) on external applied electrical field, which named as RS phenomenon [33]. Owing to the different working mechanism, ReRAMs possess the merits such as, considerable stability, ultrahigh packaging density, capacity for high speed operation, extremely low power consumption and

compatibility with standard CMOS technology [39-42], which make it a competitive candidate for the next generation of memory devices.

The memory effect of metal oxide could be traced back to 1971 when Chua described the resistive switching behavior and put forward the concept of "the Fourth device" [43]. He raised the idea that this novel device could have a hysteretic response to the applied voltage, which is the prototype of ReRAM. However, not until the year 2000 when the first feasible scheme of ReRAM was suggested by Beck et al. [44]. In 2005, the first TiO_2 based non-volatile memory was fabricated by Choi et al. [45] and lead to much more efforts were put into the development of ReRAM. In 2008, D. Strukov clearly built the connection between Chua's memristor model and the nano-scale hysteretic current behaviour [45]. After these pioneering work, thousands of breakthrough in this field have been made to obtain the optimal materials, structure and improve device performance [32, 46, 47].

The ReRAM device structure involves a functional oxide layer sandwiched by two metal electrodes, which is usually called Metal/Oxide/Metal (MOM) structure (also referred as MIM in some literatures), as shown in **Fig** 1.14. The oxide materials adopted in ReRAM devices could be Perovskite or transition metal binary oxide like TiO_2, HfO_2, NiO, ZrO_2, $SiTiO_3$ [40, 48]. Each of them has its own benefits and drawbacks. However, the recent research indicates that TiO_2 is the preferable oxide in ReRAM due to its ease of fabrication, high K, stability and exhibits both unipolar and bipolar RS phenomenon [40]. The electrode metal should fit the oxide material to obtain an overall performance. The most widely used electrode materials include Ti, TiN, Al, Pt [39]. Due to its simple structure, multilevel stacking device is also proposed to improve the overall performance [39, 49].

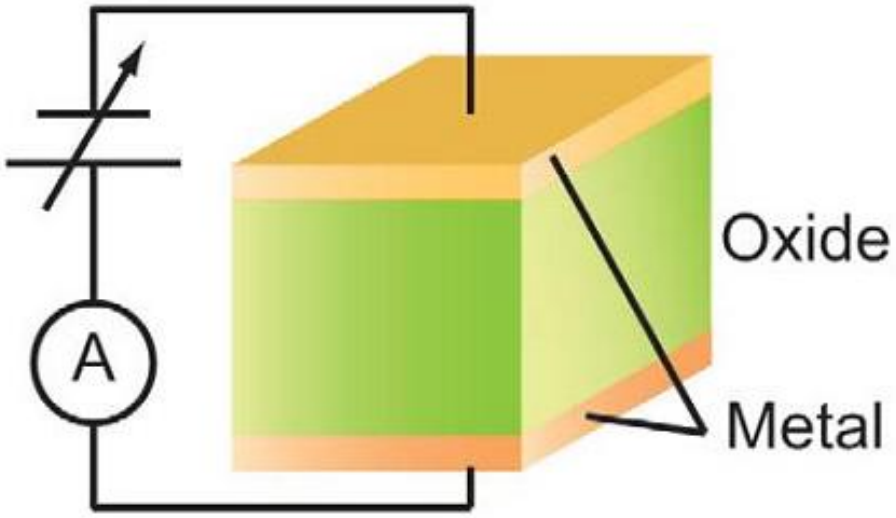


Fig 1.14 Schematic diagram of a ReRAM cell with a capacitor-like structure. Figure reproduced from Ref[51]

In a ReRAM device, the conductivity is modulated by the external applied electric field. The device may show a high resistance state (HRS)/OFF state, or a low resistance state(LRS)/ON, determined by the history of applied voltage [48]. The two different conductivity states could be defined as "0" and "1", thus makes the memory. The RS phenomenon could be classified into two types, the unipolar resistive switching and bipolar resistive switching, according to the I-V curve[49], see **Fig** 1.15. In the unipolar resistive switching, the conductivity is modulated by the amplitude of the applied voltage. The HRS changes to LRS when a high voltage (3.3eV here) is applied, which is called SET process. The LRS comes back to the HRS when a small sweeping voltage in the same direction is applied, which is called RESET process. On the contrary, the bipolar resistive switching depends on the polarity of the applied voltage, so SET and RESET happens when large enough positive or negative voltage is applied. The as-prepared device usually exhibits a high resistance and a large enough voltage is required to change to the LRS for the first time, called FORM process. In addition, to "read" the data stored in this kind of memory device is accomplished by testing the resistance with a small voltage that will not change the state. Most ReRAM devices utilize RS phenomenon in this way.

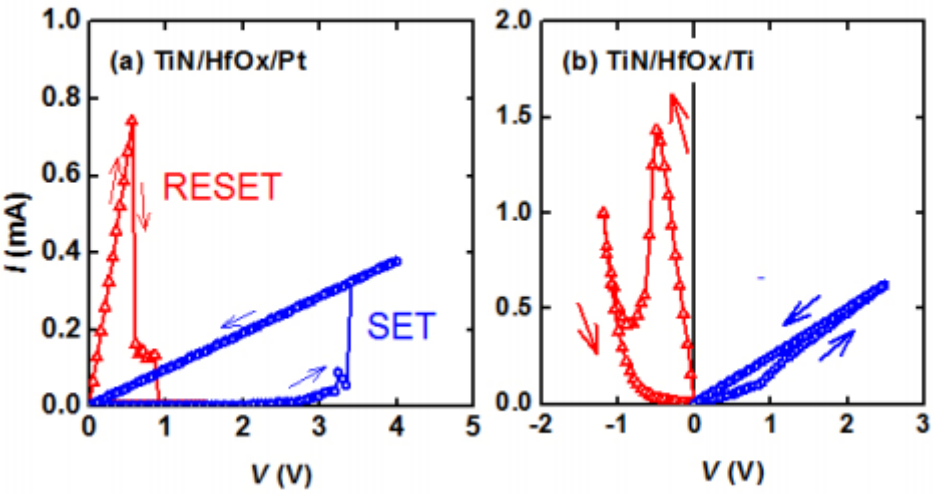


Fig 1.15 I-V characteristics of two operation modes of ReRAM (a) Unipolar and (b) bipolar modes. The arrows show the current sweep direction. Set and Reset are shown with blue and red lines separately. Figure reproduced from Ref[39]

The two popular conducting models explain the RS phenomenon best at current stage [49], although the underlying mechanism of RS is still under discussion, and not clearly illustrated. The first model indicates a conducting filament inside the oxide, and the resistance switching is explained by the forming and rupture of this conducting filament, see **Fig** 1.16. For the TiO_2 based ReRAM, filament may be associated with oxygen vacancies and titanium interstitials, and these charged defects could be driven by the external electric field to form and rupture the filament. In the year 2010, Deok-Hwang et al. have shown that the switching process is related to the forming and disruption of Ti_nO_{2n-1}(called Magneli phase) [32], which is confirmed by the high-resolution transmission electron microscopy (HRTEM) image. According to their study, oxygen vacancies could be created randomly at initial stage by thermal effect or external field, and these vacancies tend to form an ordered structure. Seong-Geon et al. also gave the DFT simulation results that oxygen vacancies ordering in certain direction is thermodynamically favourable [50]. Both experimental and theoretical research indicates that oxygen vacancies play an important role in the “filament” conducting model.

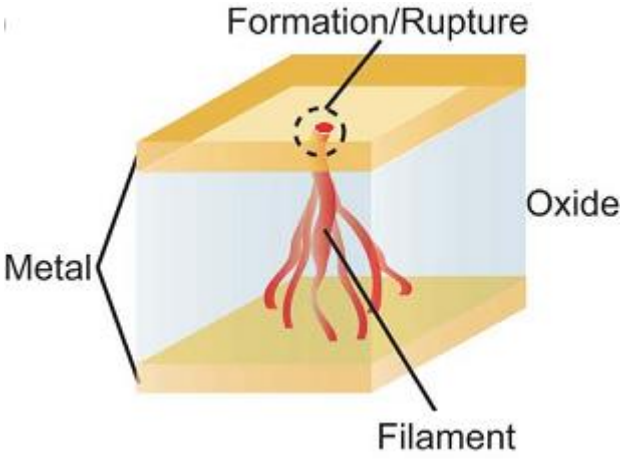


Fig 1.16 Filament model for R`eRAM operation. Figure reproduced from Ref [49]

In the second model of RS mechanism, the interface between electrode metal and oxide acts as the conducting path, as shown in **Fig** 1.17. Baikalov et al. have shown that the contact resistance of metal/oxide interface relies on the applied voltage [51]. Yang et al. provided the experimental evidence that the drift of charged oxygen vacancies by external field could lead to electronic barrier at the interface [52]. The conducting channel was switched ON when oxygen vacancies move towards the interface, and the conducting channel was switched OFF when these charged defects move away from the interface. This "interface" model usually explains the Perovskite type ReRAM well, and the driving mechanism is usually attributed to the electrochemical migration of oxygen vacancies.

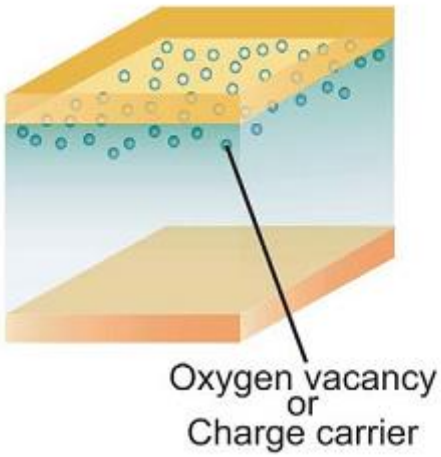


Fig 1.17 Interface model for ReRAM operation. Figure reproduced from Ref[49]

Considering the both models of RS mechanism, it is well illustrated that the defects in oxide, especially oxygen vacancies, are closely related to the change of conductivity in ReRAM. However, the clear picture of conducting channel is still unavailable, and the formation as well

as the migration mechanism of these defect are still unclear either in experiment or in theoretical research. Thus, a clear description of oxygen vacancies behaviours and properties in transition metal oxide is in urgent need.

7. IGZO Instability

As the response to the requirement for modern high-resolution display device, tremendous efforts have been made in this field for the next generation of display applications. However, a-Si suffers from poor field effect mobility, which is usually lower than 1 cm^2/Vs [53,54]. Amorphous oxide semiconductors (AOS) has been applied to thin film transistors(TFT) to substitute a-Si, for their high mobility, low fabrication temperature such as room temperature sputtering, as well as the realization of uniform properties across large device area. AOS TFT has been widely used in including the active matrix liquid crystal display (AMLCD) and also as the driver transistors of the organic light emitting diode (OLED).

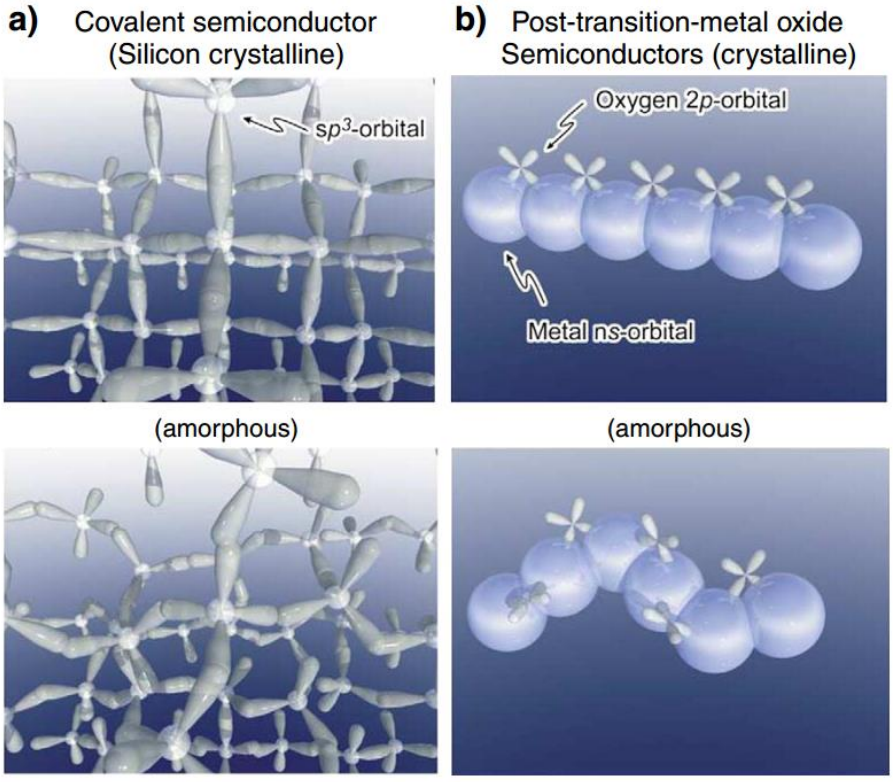


Fig 1.18 Schematic orbital structure of the conduction-band minimum in Si and in an ionic oxide semiconductor. Figure reproduced from Ref [53].

The first research on ZnO-based TFT device was published in 1968 [55] (however in crystal phase), after which the ZnO TFT device has been further developed for higher electron mobility [56] and lower turn on voltage[57] for practical use purpose. However, it was found that

polycrystalline ZnO structure was usually formed in fabrication instead of single crystal or amorphous. The existence of grain boundary causes non-uniform performance of across different location of the device, which remain to be a great challenge to single flat panel display.

The pioneering work of Nomura et al. on amorphous In-Ga-Zn-O (IGZO) TFT has provided a good solution to this problem[58]. IGZO could be easily made amorphous via room temperature deposition, and it also show high field effect electron mobility of over 5 cm^2/Vs[59-61]. The origin of high mobility in such amorphous structure has been attributed to the direct overlap between neighbouring metal s orbital [53, 62], as shown in **Fig** 1.18.

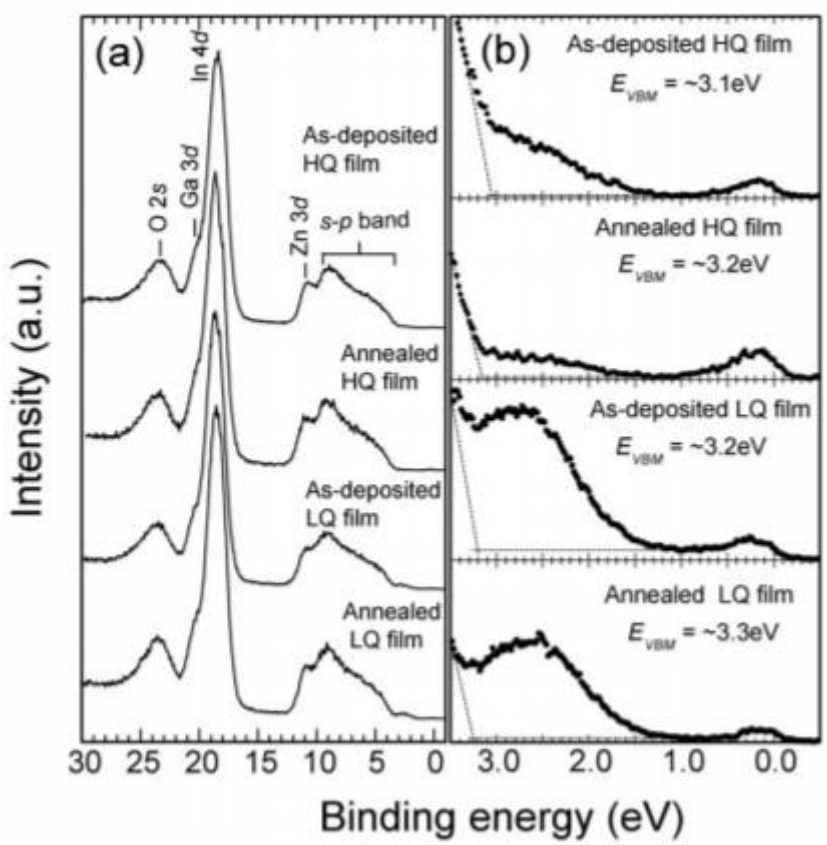


Fig 1.19 HX-PES spectra. (a) VB spectra for the a-IGZO films. (b) A magnified view around the bandgap. Intensities are normalized with the s-p band observed in (a). Figure reproduced from Ref [65]

However, there are still stability problem within a-IGZO TFT device, which usually appears as the variation in the threshold voltage V_{th} during the operation of the device. The most important stability problem is the negative bias illumination stress (NBIS) instability, which is related to the instability under the condition of light illumination and negative bias. This is because the TFTs are exposed to illumination from underlying backlight during device

operation, and the susceptibility of TFT under illumination should thus be minimized as much as possible. In addition, the device unit experience the negative bias to maintain the 'OFF' state at the most time. Therefore, the NBIS instability is the most important concern because any small change in V_{th} at such condition will make the pixels not maintain 'OFF' state as expected and blur the image.

A typical origin for the NBIS instability is the defect related photo-generated charge carriers. The sub-bandgap light excites electrons from some filled defect states lie just above valence band edge, and produce the persistent photo-conductivity[63,64]. The photo-current shifts the V_{th}, and causes the instability. The defect states responsible are seen just above the valence band edge via hard X-ray photo-emission[65], as shown in **Fig** 1.19, while the cause of these defect states are still unknown. In addition, it is also found that the instability becomes worse for more O-deficient environment[66,67], and higher humidity condition[68], as shown in **Fig** 1.20.

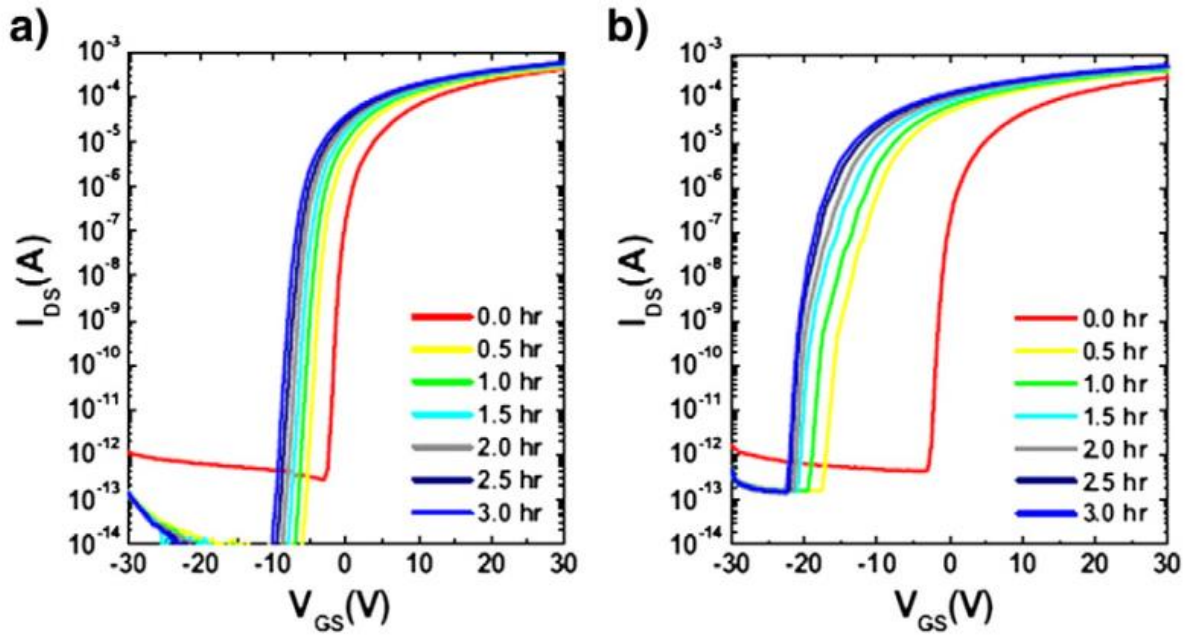


Fig 1.20 The evolution of the transfer curves as a function of the applied NBIS time at relative humidities of (a) 0% and (b) 70%, respectively. Figure reproduced from Ref [68].

There have been various models to account for this instability, including oxygen vacancies [64,66,69-71], oxygen interstitials [72-75], or hydrogen impurities [73,76,77], but each of these models has its flaws. The O vacancy states lie too high in energy in the ZnO band gap [78,79] and IGZO (unpublished, in preparation) compared to their energy as seen by photoemission[65]. The second model based on oxygen interstitials does produce states in the required energy

range [73]. However, the problem is that they are an oxygen excess defect while annealing the film in an oxygen excess is known to reduce the instability [66,67]. A final possibility is that it is due to hydrogen. It has been proposed that single hydrogen atoms could undergo a deep to shallow transition and become donor-like [76,77]. However, this defect does not correspond to an oxygen deficiency and there is little other evidence for this mechanism.

Therefore, it is desired to understand the origin of NBIS instability, especially the actual defect related to it. The defect should be O deficient defect with states just above valence band edge. It is also worthwhile to investigate the effect of water or hydrogen in this mechanism. The chapter 7 of this thesis will put forward a new mechanism to explain the NBIS instability mechanism in a-IGZO, which meets all these requirements.

8. Thesis Aim and Outline

The aim of this thesis is to investigate underlying physics mechanism of existing problems in MOSFET and ReRAM which are related to defects in oxides. The density functional theory simulation is carried out to provide a clear theoretical explanation in atomic scale. The whole thesis is organized as follows:

Chapter 1 summarizes the research history of Ge MOSFET and ReRAM, including bottleneck Si MOSFET scaling, opportunities for Ge as high mobility channel in MOSFET, challenges of oxides to be compatible with Ge channel in MOSFET, and defect related working mechanism of ReRAM.

Chapter 2 presents the theoretical background for the simulation method applied in this study. It covers quantum mechanics, electron-ion interaction, pseudopotential supercell approach, density functional theory, hybrid functional, and the CASTEP software package itself.

Chapter 3 researches on the rare earth element doping in GeO_2. Amorphous atomic structure of doped GeO_2 network and Ge:GeO_2 interface are built to investigate the improved device reliability observed in experiments. Various O deficiency defects in different chemical environment are analysed and the improved reliability are attributed to the passivation of VAP type O deficiency defects by doped rare earth elements.

Chapter 4 investigates the GeO_2 oxidation mechanism. Both O excess and O deficiency defects' formation and diffusion behaviours in SiO_2 and GeO_2 are simulated. The type of defect with lowest average migration barrier are found to be different in SiO_2 and GeO_2. The GeO_2 oxidation mechanism is put forward to be via lattice O interstitial diffusion, compared to O_2

molecule diffusion in SiO_2. This new mechanism fully explained the strange isotope tracer experiment results in literature [25].

Chapter 5 researches on the metal semiconductor electrical contact involved in Ge MOSFET. Compared to the strong Fermi level pinning of normal metal on top of Ge, the Fermi level pinning are proved to be much weaker in Ge:germanide interface. Several new Ge:Gemanide interface models with little lattice stress are put forward in this study. The weak Fermi level pinning of germanide results are in good consistence with experiments and well explained by the interfacial states pinning. These results are important to tune SBHs for n-type contacts on Ge for use of Ge high mobility substrates in future CMOS devices.

Chapter 6 investigates the surface and subsurface O vacancy defects in three kinds of stable TiO_2 surface. The charge transition levels are calculated for O vacancy defects on these surfaces. The low formation energy at O poor condition and the +2 charge state as the most stable state of O vacancy is beneficial to the formation and rupture of conducting filament in ReRAM, which makes TiO_2 a good candidate for ReRAM material.

Chapter 7 investigates the hydrogen behaviours in amorphous ZnO. It is found that H exists as hydrogen pairs trapped at oxygen vacancies and form Zn-H bonds. This is different from crystal ZnO situation where H acts as shallow donor. The O vacancy 2H complex defect has got defect states in lower gap region, which is proposed to be origin of the negative bias light induced stress instability.

Chapter 8 concludes the work and suggests further investigations.

9. Reference

[1] U.S. Patent 3,102,230 filed in 1960, issued in 1963

[2] Robert H. Dennard, et al. IEEE Journal of Solid State Circuits. SC–9 (5), 1974

[3] H. Iwai, Microelectron. Eng. 86, 1520 (2009)

[4] J. Robertson, Rep. Prog. Phys. 69,327 (2006)

[5] E. Arnold, et al. Appl. Phys. Lett. 13, 413 (1968)

[6] T. Hattori, et al. Appl. Phys. Lett. 43, 470 (1983)

[7] F. J. Grunthaner, et al. Phys. Rev. Lett. 42, 1683 (1979)

[8] F. J. Himpsel, et al. Phys. Rev. B. 38, 6084 (1988)

[9] E. H. Poindexter, et al. J. Appl. Phys. 56, 2844 (1984)

[10] C. Kittel, Introduction to Solid State Physics, Wiley (2004)

[11] G.K. Teal, et al. Proc. I.R.E. 40, 906 (1952)

[12] N. Yeh, et al. J. Mater. Chem. C, 1, 4616 (2013)

[13] S. Takagi, Technical Digest of Int. Electron Devices Meeting. 23.1(2012)

[14] R. Zhang, et al. Jpn. J. Appl. Phys. Part 1, 54, 06FA01 (2015)

[15] A. Toriumi, et al. Technical Digest of Int. Electron Devices Meeting. 28.4 (2011)

[16] T. Sasada, et al. J. Appl. Phys. 106, 073716 (2009)

[17] R. Zhang, et al. Appl. Phys. Lett. 102, 081603 (2013)

[18] A. Delabie, el al. Appl. Phys. Lett. 91, 082904 (2007)

[19] R. Zhang, et al. Appl. Phys. Lett. 98, 112902 (2011)

[20] Y. Kamata, et al. Jpn. J. Appl. Phys. 45, 5651 (2006)

[21] Toriumi A et al. 2011, 35(3): 443-456.

[22] Zhang, Rui, et al. Electron Devices Meeting (IEDM), 2011 IEEE International. IEEE, 2011.

[23] L. Skuja, J. Non-Cryst Solids 149, 77 (1992)

[24] M. Fujimaki et al., Phys. Rev. B 57, 3920 (1998)

[25] S K Wang, et al. Jpn J App Phys **50** 04DA01 (2011)

[26] C. Lu et al in Technical Digest of IEDM, p. 14.6 (2015)

[27] C. H. Lee, et al. APEX 5 114001 (2012)

[28] Lin, Dennis, et al. Inter. Electron Devices Meeting (IEDM), 2012

[29] Heine, V. Phys. Rev. **138**, 1689 (1965).

[30] Tejedor, C. et al. J. Phys. C: Solid State Phys. **10**, 2163 (1977).

[31] Tejedor, C. et al. J. Phys. C **11**, L19 (1978)

[32] Kwon, D.H. et al. Nature Nanotechnology, 2010. 5(2): p. 148-153.

[33] Zhu, et al. Frontiers of Materials Science, 2012. 6(3): p. 183-206.

[34] Choi, B.J., et al. Journal of Applied Physics, 2005. 98(3).

[35] Fujimoto, M., et al. Applied Physics Letters, 2006. 89(22).

[36] Song, S.J. et al. Scientific Reports, 2013. 3.

[37] Liu, Q. et al. ACS Nano, 2010. 4(10): p. 6162-6168.

[38] Zhao, L. et al. Nanoscale, 2014. 6(11): p. 5698-5702.

[39] Akinaga, et al. IEICE ELECTRONICS EXPRESS, 2012. 9: p. 795-807.

[40] Acharyya, D. et al. Microelectronics Reliability, 2014. 54(3): p. 541-560.

[41] Tian, X.Z., et al. Science China-Physics Mechanics & Astronomy, 2013. 56(12): p. 2361-2369.

[42] Zhu, X.J. et al. Chinese Science Bulletin, 2014. 59(20): p. 2363-2382.

[43] Chua, L.O. IEEE Transactions on Circuit Theory, 1971. Ct18(5): p. 507-&.

[44] Beck, A, et al. Applied Physics Letters, 2000. 77(1): p. 139-141.

[45] Strukov, D.B. et al. Nature, 2008. 453(7191): p. 80-83.

[46] Shin, J., et al. Journal of Applied Physics, 2011. 109(3).

[47] Jeong, H.Y. et al. Nanotechnology, 2010. 21(11).

[48] Lanza, M. Materials, 2014. 7(3): p. 2155-2182.

[49] Sawa, A. Materials Today, 2008. 11(6): p. 28-36.

[50] Park, S.G. et al. IEEE Electron Device Letters, 2011. 32(2): p. 197-199.

[51] Baikalov, A. et al. Applied Physics Letters, 2003. 83(5): p. 957-959.

[52] Yang, J.J. et al. Nature Nanotechnology, 2008. 3(7): p. 429-433.

[53] M.J. Powell, Appl. Phys. Lett. 43 (1983) 597

[54] F.R. Libsch, J. Kanicki, Appl. Phys. Lett. 62 (1993) 1286

[55] G.F. Boesen, J.E. Jacobs, Proc. IEEE 56/11 (1968) 2094

[56] L. Zhang, J. Li, X.W. Zhang, X.Y. Jiang, Z.L. Zhang, Appl. Phys. Lett. 95 (2009) 072112.

[57] S.J. Lim, S.J. Kwon, H. Kim, J.S. Park, Appl. Phys. Lett. 91/18 (2007)

[58] K. Nomura, H. Ohta, A. Takagi, T. Kamiya, M. Hirano, H. Hosono, Nature 432(2004) 488

[59] Hosono, H. J. Non. Cryst. Solids 352, 851 (2006)

[60] Kamiya, T. & Hosono, H. NPG Asia Mater 2, 15 (2010).

[61] Fortunato, E., Barquinha, P. & Martins, R. Adv Mater 24, 2945–2986 (2010)

[62] Park, J. S., Maeng, W. S., Kim, H. S. & Park, J. S. Thin Solid Films 520, 1679 (2012).

[63] Ghafarzadeh, K. et al. Appl. Phys. Lett. 97, 113504 (2010)

[64] Jeon, S. et al. Nature Mater. 11, 301 (2012).

[65] Nomura, K. et al. Appl. Phys. Lett. 92, 202117 (2008).

[66] Ji, K. H. et al. Appl. Phys. Lett. 98, 103509 (2011).

[67] Yang, S. Appl. Phys. Lett. 99, 102103 (2011).

[68] K.-H. Lee, et al. Appl. Phys. Lett. 95 (2009) 232106

[69] Ryu, B. et al. Appl. Phys.Lett 97, 022108 (2010).

[70] No, H. K. et al. Phys Rev B 84, 1115205 (2011).

[71] Yang, S. et al. Appl. Phys. Lett. 99, 102103 (2011).

[72] Nahm, H. et al. Phys Stat Solidi B 249, 1277 (2012).

[73] Robertson, J. & Guo, Y. Appl. Phys. Lett. 104,162102 (2014).

[74] Walsh, A., et al. Chem Mater 21, 5119 (2009).

[75] Sallis, S. et al. Phys. Stat. Solidi.A 212, 1461 (2015).

[76] Nahm, H.,et al. Sci. Reports 4, 4124 (2014).

[77]. Kang, Y. et al. Adv. Electronic Mater. 1, 1400006 (2015)

[78] Janotti, A. & de Walle, V. C.G. Appl. Phys. Lett. 87, 122102 (2005).

[79] Oba, F., etl al. Phys. Rev.B 77, 245202 (2008)

Chapter 2 Theoretical Simulation Methods

1. Introduction

Theoretical simulation based on quantum mechanics have shown its powerful ability to predict and explain the fundamental properties of materials including metal, metal oxide, organic materials, etc. By calculating the interactions between electrons-electrons, electrons-ions and ions-ions, we could obtain the electronic parameters of the system, such as band gaps, defect states, mobility, and further analyze the device performance.

Currently, the most popular quantum mechanics simulation method is the ab initio methods, or the first principles simulation. The ab initio method is the calculation based purely on the information of atomic type and location in the system without any empirical model or fitting parameters. It usually determines the ground state by minimizing the total energy of the whole system and then gives the materials properties based on the convergent electronic density distribution. Nearly every microscopic property could be simulated in this way. The molecular configuration could be extracted from the ground state, and the heat of reaction could be obtained by calculating the difference of the total energy between reactants and products. As the computation ability increases and more efficient approximation methods appear, ab initio methods could be used to simulate many large systems.

2. The Schrodinger Equation and Born-Oppenheimer approximation

For a system consists of n electrons and μ nuclei, the Schrodinger equation could be written as:

$$H\psi(r,R) = E^{H}\psi(r,R) \tag{2.1}$$

with E as the total energy of the system, H as the Hamiltonian operator and ψ as the many-body wavefunction, r and R as the coordinates for electrons and nuclei respectively.

$$H = V_{ext} + T + V_{ee} \tag{2.2}$$

In the formula above, V_{ext} refers to the potential related to external field, T refers to the kinetic energy, and V_{ee} refers to the potential associated with interactions between electrons.

In principle, the Hamiltonian operator could be divided into three part: the electrons-electrons, the nuclei-nuclei and the electrons-nuclei interaction.

$$H(r,R) = H_e(r) + H_N(R) + H_{e-N}(r,R) \tag{2.3}$$

Since the nucleus mass is much larger than the mass of the electron, the movement of nuclei is much slower than that of electrons. Thus, the state of electrons changes quickly while nuclei only vibrate slowly around their balanced positions. In principle, we could consider the movement of electrons and nuclei separately, that nuclei are assumed to stay static when be calculate movement of electrons, and when the nuclei change their configuration, electrons can respond in time and come to a stable distribution.

$$\psi_{n\mu}(r,R) = \chi_{n\mu}(R)\phi_n(r,R) \tag{2.4}$$

In the formula above, $\chi_{n\mu}(R)$ describes the wavefunction of nuclei and $\phi_n(r,R)$ describes the wavefunction of electrons, which should be anti-symmetric according to Pauli Exclusion Principle. That's the adiabatic approximation. Usually, the kinetic energy of nuclei is neglected since they are assumed to be static when calculate the movements of electrons, and the ions-ions interaction term is also neglected for the same reason. Thus the Hamiltonian operator could be expressed as following equation (the atomic unit applies to all of equations in this chapter):

$$H = -\frac{1}{2}\sum_i \nabla_{r_i}^2 + \sum_i V(r_i) + \frac{1}{2}\sum_{i\neq i'} \frac{1}{|r_i - r_{i'}|} \tag{2.5}$$

This is the first step of Born-Oppenheimer approximation, in which we can obtain the electron related energy E_e as the function of nuclear coordinates R, which is also known as potential energy surface. In the second step, the nuclei related energy is reintroduced. The two steps jointly forms the Born-Oppenheimer approximation.

3. Hartree-Fock Approximation

Through the Born-Oppenheimer approximation, we separate the movement of electrons and nuclei, which leads to the multi-electrons Schrodinger equation,

$$\left[-\frac{1}{2}\sum_i \nabla_{r_i}^2 + \sum_i V(r_i) + \frac{1}{2}\sum_{i,i'} \frac{1}{|r_i - r_{i'}|} \right] \phi = E\phi \tag{2.6}$$

The Hartree-Fock approximation introduced here neglect the interactions between electrons, treats system as a non-interating system with only the exchange property. At this approximation,

the multi-electrons wavefunction could be written in the form of product of single-electron wavefunctions.

$$\phi(r) = \varphi_1(r_1)\varphi_2(r_2)\cdots\varphi_n(r_n) \tag{2.7}$$

$$\left[-\frac{1}{2}\nabla^2 + V(r) + \sum_{i'(\neq i)}\int dr' \frac{\left|\varphi_{i'}(r')\right|^2}{\left|r'-r\right|}\right]\varphi_i(r) = E_i\varphi_i(r) \tag{2.8}$$

The formula above is called the Hartree wavefunction, and in this way, we only need to deal with the single-electron wavefunction although we are simulating the many-body system. However, the multi-electron wavefunction obtained by the product of single-electron wavefunction should also change sign when exchanging two electrons that have the same spin according to the Pauli Exclusion Principle. To solve this problem, the Fock approximation is introduced. The wavefunctions are written in a Slater determinant:

$$\varphi^{HF}(\{\mathrm{r}_i\}) = \frac{1}{\sqrt{n!}}\begin{vmatrix} \varphi_1(r_1) & \cdots & \varphi_1(r_n) \\ \cdots & \cdots & \cdots \\ \varphi_n(r_1) & \cdots & \varphi_n(r_n) \end{vmatrix} \tag{2.9}$$

Where n is the number of electrons and $\varphi(r_1)$ is the normalized single particle state. We use this Slater determinant to replace the multi-electron wavefunction, and come to the Hartree-Fock equation:

$$\left[-\frac{1}{2}\nabla^2 + V(r) + \sum_{i'(\neq i)}\int dr' \frac{\left|\varphi_{i'}(r')\right|^2}{\left|r'-r\right|}\right]\varphi_i(r) - \sum_{i'(\neq i),\|}\int dr' \frac{\varphi_{i'}^*(r')\varphi_i(r')}{\left|r-r'\right|}\varphi_{i'}(r) = E_i\varphi_i(r) \tag{2.10}$$

Thus, the original many body problem evolves to one set of single electron problem, and the effect of the other n-1 electrons could be seen as the mean field applied on the single electron. As discussed earlier, Hartree-Fock approximation totally neglects the correlation properties, and can give results quite different from experiment. It usually overestimate the band gap of semiconductor and insulators due to the strong correlation in solid. In addition, Hartree-Fock calculation is slow, thus more advanced techniques are needed to speed up the calculations.

4. Hohenberg-Kohn Theory and DFT

The computational cost of Hartree-Fock method scales as n^4, which increase rapidly as the number of electrons increase. Thus, simulation based on Hartree-Fock could only be used to deal with very small systems. An alternative method is to describe the system with electron density ρ instead of calculation of multi-electron wavefunction, which is based on the two remarable statements given by Hohenberg and Kohn [1].

(1) The external potential $V_{ext}(r)$ is uniquely determined by the corresponding ground-state electronic density, to within an additive constant.
If two systems of electrons trapped at two different external potential ν_1 and ν_2 have the same ground state density, then necessarily,

$$\nu_1 - \nu_2 = \text{const.} \tag{2.11}$$

(2) For all ν-representable densities $\rho(r)$, $E_V[\rho] \geq E_0$ where E_0 is now the ground-state energy for n electrons in the external potential $V_{ext}(r)$.

The total energy could be rewritten as the form of:

$$E = \langle \varphi | H | \varphi \rangle = \int V_{ext}(r)\rho(r)d(r) + \langle \varphi | T + V_{ee} | \varphi \rangle \tag{2.12}$$

Hohenberg-Kohn Theory showed that there is a one-to-one relationship between the density ρ and the external potential, thus total energy E is a functional of the electronic density of the ground state for a given external potential ν.

$$E_\nu = \int V_{ext}(r)\rho(r)d(r) + \langle \varphi | T + V_{ee} | \varphi \rangle = \int V_{ext}(r)\rho(r)d(r) + \mathrm{F}(\rho) \tag{2.13}$$

The $F(\rho)$ here is a universal functional which does not rely on the external potential. The Density Functional Theory (DFT) is based on such theorem, describe the system with electronic density distribution, which save a lot of computational resources compared to Hartree-Fock methods.

5. The Kohn-Sham Procedure

The Hohenberg-Kohn Theory shows that the one-to-one relationship between the electronic density and the external field. However, it gives no details on how to construct the functional $F(\rho)$. The Kohn-Sham procedure is a method to find the universal functional $F(\rho)$, which assume that the kinetic energy functional of the real system could be replaced by that of a non-

interacting system [2]. The electronic density is constructed as the sum of n single-electron wavefunction:

$$\rho(r) = \sum_{i=1}^{N} \left|\varphi_i(r)\right|^2 \tag{2.14}$$

This density should obey the law that the total number of particle is unchaneged:

$$\int dr \delta\rho(r) = 0 \tag{2.15}$$

The Kohn-Sham equation could be written as:

$$\{-\nabla^2 + V_{KS}[\rho(r)]\}\varphi_i(r) = E_i\varphi(r) \tag{2.16}$$

$$V_{KS}[\rho(r)] = \upsilon(r) + \int dr' \frac{\rho(r')}{\left|r - r'\right|} + \frac{\delta E_{xc}[\rho]}{\delta\rho(r)} \tag{2.17}$$

Through the Kohn-Sham procedure, we could replace the interacting many-body problems with non-interacting electrons immersed in an effective potential applied by all the other electrons. Different approximations could be applied to construct the exchange-correlation energy E_{XC} as a functional of electron density ρ (r).

6. Exchange-Correlation Functionals

The exchange-correlation term $E_{XC}(\rho)$ is traditionally divided into the exchange term $E_X(\rho)$ and the correlation term $E_C(\rho)$. The Hartree-Fock exchange is often assumed to be the exact exchange in DFT, and the residual in $E_{XC}(\rho)$ is referred as the correlation term.

The second order density matrix in terms of the wavefunction could be defined as:

$$\rho_2(r,r') = n(n-1)\int d^3r_3 \int d^3r_n \left|\varphi(r,r',r_3,...,r_n)\right|^2 \tag{2.18}$$

Thus, the probability of finding an electron in volume d^3r at r and another electron in volume d^3r' at r' could be expressed as $\rho_2(r,r')d^3rd^3r'$. We could further define a conditional probability density:

$$\rho_2(r,r') = n(r)n_2(r,r') \tag{2.19}$$

So the $n_2(r,r')d^3r'$ is the probability of finding an electron in volume d^3r' at r', at the condition that there is an electron in volume d^3r at r. Based on this definition, the exchange correlation hole $n_{XC}(r,r')$ could be expressed as

$$n_2(r,r') = n(r') + n_{XC}(r,r') \tag{2.20}$$

So the exchange correlation hole is the difference between the probability of "finding an electron in volume d^3r' at r', at the condition that there is an electron in volume d^3r at r" and the probability of "finding an electron in volume d^3r' at r'". Physically, it describe the phenomenon that an existing electron d^3r at r will lower the probability to find another electron in volume d^3r' at r', which seems like that the first existing electron create a hole in the space whose distribution could be described by $n_{XC}(r,r')$. Through using the Hellmann-Feynman theorem [3-6], we could write the exchange-correlation energy as the electrostatic interaction between the density and the hole [7]:

$$E_{XC} = \frac{1}{2}\int n(r)d^3r \int \frac{\bar{n}_{XC}(r,r+u)}{u} d^3u \tag{2.21}$$

with

$$\bar{n}_{XC}(r,r') = \int_0^1 n_{XC,\lambda}(r,r')d\lambda \tag{2.22}$$

The basic idea here is that while keeping the density fixed, the non-interacting system is connected to the interacting system via a coupling-constant λ, which appears in the equation (2.22). λ represents the strength of the electron-electron interaction. λ = 0 implies non-interacting system, while λ = 1 implies fully interacting system.

6.1 LDA

Density functional theory provides us a powerful method to use the exchange-correlation energy functional to calculate the total energy. The simplest approximation of the exchange-correlation functional is the local density approximation (LDA), which also gives good results on most properties including bond length, configuration, and heat of reaction. In LDA, it is assumed that the exchange-correlation energy per electron at r in the electron gas, $\varepsilon_{XC}(r)$, is the same as the exchange-correlation energy per electron in a homogeneous electron system that has the same electron density as the electron gas at r. Thus, the exchange-correlation energy

$\varepsilon_{XC}(r)$ in LDA only depend on the electron density at location r, and could be approximated to be the value in homogeneous electron gas with the same electron density [8].

$$E_{XC}[n(r)] = \int \varepsilon_{XC}(r) n(r) d^3 r = \int \varepsilon_{XC}^{\text{hom}}[n(r)] n(r) d^3 r \quad (2.23)$$

We could see that LDA is only accurate for the uniform system but it is also a good approximation for the system whose electron density varies quite slowly. LDA has shown very good results (especially structural, electronic and vibrational properties) which agree well with experimental value in many fields. Considering its simplicity, LDA has been widely used to save the computing resources on condition of enough accuracy.

There are also some limitations of LDA. The binding energies given by LDA are often too large, which is often referred as the overbinding effect. The activation energies in chemical reactions suggested by LDA are also too small, such as the H_2 molecule on Al and Cu surfaces. The most important problem of LDA is the well-known band gap error, the band gap of semiconductor could be underestimated up to 50%.

6.2 GGA

In order to improve LDA's shortcomings mentioned above, the more accurate generalized gradient approximation (GGA) was proposed. The exchange-correlation functional in GGA depends not only on the local electron density but also on its gradient at the position r, which could be written in the form of:

$$E_{XC}^{GGA}[n(\mathrm{r})] = \int \varepsilon_{XC}[n(r), \nabla n(r)] n(r) dr \quad (2.24)$$

The $E_{XC}^{GGA}[n(\mathrm{r})]$ term is the exchange-correlation energy functional, $n(r)$ is the electron density and $\nabla n(r)$ is the gradient of electron density. According to this idea, many GGA functionals have been put forward, among which the PW91 form and Perdew, Burke and Enzerhof (PBE) form are most widely used.

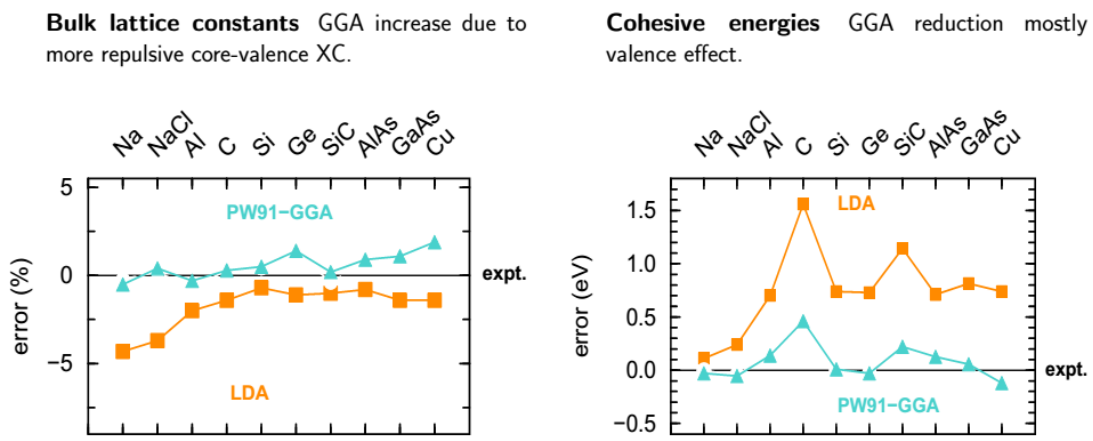


Fig 2.1 Comparison of GGA and LDA simulation results of lattice constants and cohesive energies with experimental value. Figure reproduced from Ref[14]

The GGA functional gives a better results in many fields than the LDA functional does (see **Fig** 2.1), which results from its inclusion of the electron density gradient term. The problem of too short bond lengths and lattice constants in LDA is solve to some extent in GGA, due to the more repulsive core-valence exchange-correlation interaction. The total energies and the cohesive energies are also improved thanks to the more accurate approximation. However, the band gap problem remains unsolved in GGA due to several reasons. Even if GGA has included the gradient of electron density, it is still a semi-local functional. The semi-local functional has a spurious self-interaction in the occupied states, which makes electron orbitals too delocalized and push the occupied states up in energy, thus reduce the band gap. Thus, it is not surprising that GGA cannot describe the non-local exchange-correlation interaction well.

6.3 Hybrid functional

As stated above, LDA and GGA usually underestimate the band gap of semiconductor and insulator make electronic states delocalized, while Hartree-Fock functional overestimates the band gap and makes electronic states overly localized. A natural idea is that, a careful mixing of HF and LDA/GGA may give the correct band gap and localization, which gives birth to the hybrid functional.

$$E_{XC}^{hybrid}(\varphi(r)) = \alpha E_X^{HF}(\varphi(r)) + (1-\alpha)E_X^{local}(n(r)) + E_C^{local}(n(r)) \tag{2.25}$$

where α is the fraction coefficient of Hartree-Fock exchange.

6.3.1 Heyd-Scuseria-Ernzerhof (HSE) Functional

HSE functional mix the functional depending on the distance. The PBE0 functional mixes the

PBE exchange energy with exact nonlocal Fock exchange energy. Considering that the long-range Fock exchange is time consuming, HSE retains only the short-range part of Fock exchange while preserve accuracy [9]. The long-range and the remaining short-range exchange are extracted from PBE functional.

$$E_{XC}^{HSE} = \alpha E_X^{HF,SR}(\omega) + (1-\alpha)E_X^{PBE,SR} + E_X^{PBE,LR} + E_C^{PBE} \tag{2.26}$$

where α =0.25 for HSE. This functional has been proved to be timesaving and more accurate compared with PBE0.

6.3.2 Screened Exchange-LDA Functional (sX-LDA)

Bylander and Kleinman put forward the concept of screened exchange (sX), which combines a short-range fraction of Fock exchange with a semi-local model of long-range exchange [10]. This idea is applied in the sX-LDA functional, in which the exchange term is divided into the Thomas-Fermi screened exchange part and the residual, which is often referred as the Thomas-Fermi model of screening [11-13]. The screened exchange energy could be expressed as:

$$E_X^{sX}[\varphi] = -\sum_{i<j}^{N} drdr' \times \frac{\varphi_i^*(r)\varphi_j^*(r')e^{-k_{TF}|r-r'|}\varphi_j(r)\varphi_i(r')}{|r-r'|} \tag{2.27}$$

with k_{TF} as the Thomas-Fermi screening constant, and φ as the Kohn-Sham single particle state. Enough research work has been carried out to prove that the band gaps and electronic states calculated by sX-LDA agree well with experimental value.

7. Periodic Supercell and Plane Wave Basis Set

In order to simulate the solid state system, it is inevitable to deal with the perfect crystal, which involves infinite electrons. For such system, infinite number of wave functions and basis set are required to describe electrons' behaviors. One possible method to solve this problem is to utilize the periodicity of the system, and build the periodic supercell[14], as shown in **Fig** 2.2. Only finite number of atoms are involved in the supercell and the other atoms in the crystal are represented by the image of these atoms in the supercell. Bloch's theorem ensure the supercell system's properties are consistent with the infinite crystal system.

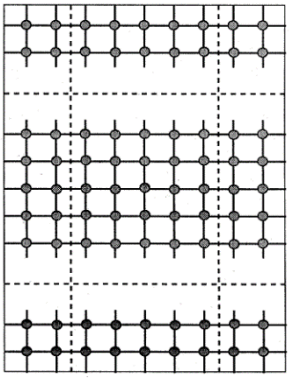


Fig 2.2 Periodic supercell of the slab model. Figure reproduced from Ref[14]

According to Bloch's theorem, the wave function of an electron in a periodical system could be expanded as the product of a periodic part and a wave part,

$$\varphi_i(r) = f_i(r)\exp(j\vec{k}\bullet\vec{r}) \tag{2.28}$$

and that the periodic part $f_i(r)$ could be expressed in the form of one set of plane waves with vectors the same as the lattice reciprocal vector :

$$f_i(r) = \sum_G c_{i,G}\exp(j\vec{G}\bullet\vec{r}) \tag{2.29}$$

Here, G is the lattice reciprocal vector. Thus, the wave function could be written in the form of:

$$\varphi_i(r,\vec{\mathrm{k}}) = \sum_G c_{i,G}\exp[j(\vec{\mathrm{k}}+\vec{\mathrm{G}})\bullet\vec{r}] \tag{2.30}$$

The above formula consists of infinite number of plane waves, in which the ones with small kinetic energy $\frac{\hbar^2}{2m}\left|\vec{k}+\vec{G}\right|^2$ are of more importance, while the high energy plane waves represent the intense variation of electron density. It is sensible to truncate the plane wave expansion at certain cut-off energy so only finite number of plane waves are involved in the expression of wave function.

The vector $\vec{k}$ should be chosen carefully in simulation. Each $\vec{k}$ could be represented by a point in the reciprocal space, which is named as k point. Each k point is associated to a plane wave, thus represent the electronic state. A useful technique is to choose some k points distributed in the reciprocal space, each one could represent all possible k values in its vicinity,

and take the weighted mean value of these chosen k points' results. This process is often referred as k-points sampling. The well-known Monkhorst-Pack sampling scheme is to choose a uniform distribution of k points in reciprocal space and only calculate the ones located in the first Brillouin zone, which utilize the periodicity of the system[15]. The density of the sampling k points really depends on the materials, and a k-points convergence test is usually the best way to ensure the sensible results.

8. Pseudopotential

The simulation based on all electrons and nuclei is usually time consuming, and the introduction of pseudopotential could greatly improve the efficiency while preserve the accuracy of simulation. Real electrons could be classified into two categories, the core electrons and the valence electrons. Valence electrons determine the main physical and chemical properties of materials since they participate in the bonding process. However, it takes much more plane wave to describe the behaviors of the core electrons, and the valence electron also oscillate greatly in the core region. In order to reduce the number needed of plane waves to describe the system, we could replace the real core electrons and nuclei with pseudoelectrons, which behave nearly the same as the real core electrons and nuclei[16-18]. In addition, real valence electrons are substituted with pseudoelectrons with less oscillation in the core region. The pseudoelectrons must behave nearly identically to the interaction between real electrons and nuclei. M. C. et al. described the process to construct pseudopotentials[14].

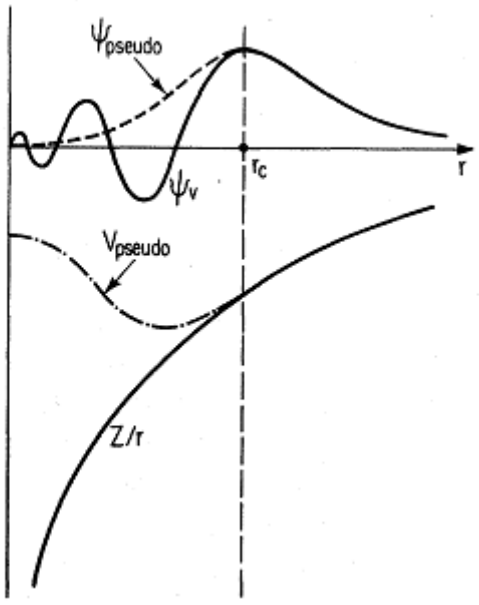


Fig 2.3 Pseudopotential and corresponding pseudo wave function [14]

There are two prevalent types of pseudopotentials, the norm-conserving ones and the ultrasoft ones. The norm-conserving requires the integrated charge is the same for all electron wave functions and pseudo wave functions. This ensures that the total charge in the core region is correct while the normalized pseudo wave function is equal to the all electron wave function ouside of cutoff radius. On the contrary, the ultrasoft pseudopotentials do not obey this condition and thus have more freedom. As the result, the ultrasoft pseudopotential has lower cutoff energy and requires less computational resources.

9. CASTEP Software Package

CASTEP is an academic software package using DFT with a plane wave basis set[19]. It treats materials as periodic supercell thus most suitable for crystal materials simulation, however it can also be used for amorphous structure, surface structure, and other situation with less symmetry. It could be applied to calculate a wide range of properties, including configuration, total energy, electronic properties, and vibrations of many materials from first principles method. Functions including single point energy, geometry relaxation, molecular dynamics, transition state search could be realized by this package. In this research, both CASTEP 8.0 and the commercial version Material Studio 7.0 are used.

10. Reference

[1] P. Hohenberg, et al Phys. Rev. B 136, 864 (1964).
[2] W. Kohn, el al. Phys. Rev 140, A1133 (1965)
[3] P. Güttinger, Z. Phys. 73, 169 (1932)
[4] W. Pauli, Handbuch der Physik 24 (Berlin:Springer)
[5] H. Hellmann, Z Phys, 85, 180 (1933)
[6] R. P. Feynman, Phys. Rev. 56, 340 (1939)
[7] R. G. Parr, et al. Density Functional Theory of Atoms and Molecules, (Oxford, New York, 1989)
[8] J. P. Perdew, et al. Phys. Rev. B 45, 13244 (1992)
[9] M. A. L.et al. Phys. Rev. B 83, 035119 (2011)
[10] D. M. Bylander, et al. Phys. Rev. B 41, 7868 (1990).
[11] J. C. Phillips, et al. Phys. Rev. 128, 2098 (1962)
[12] L. Hedin, Phys. Rev. 139, A796 (1965)
[13] S. J. Clark, et al., Phys. Rev. B 82, 085208 (2010)
[14] Payne, M.C, et al. Reviews of Modern Physics, 1992. 64(4): p. 1045-1097.

[15] H. J. Monkhorst and J. D. Pack, Phys. Rev. B 13(12), 5188 (1976)

[16] J. C. Phillips, Phys. Rev. 112, 685 (1958).

[17] M. L. Cohen and V. Heine, Solid State Physics 24, 37 (1970).

[18] M. T. Yin and M. L. Cohen, Phys. Rev. B 25, 7403 (1982)

[19] S. J. Clark, Zeitschrift fuer Kristallographie 220(5-6) pp. 567-570 (2005)

Chapter 3. Doping in GeO_2 and MOSFET Reliability

1. Overview

High mobility channel materials such as Ge, SiGe and InGaAs are of critical importance for the continued scaling of complementary metal oxide semiconductor (CMOS) transistors. Ge, as an elemental semiconductor with larger electron and hole mobilities than Si, would in principle be particularly favored. However, its native oxide GeO_2 and the Ge/GeO_2 interface suffer from worse electronic properties than the Si/SiO_2 interface [1-8], a larger interface trap density (D_{it}), worse electrical reliability [6,7] and unusual oxidation kinetics [8]. At some level, it appears that oxygen deficiency defects form in GeO_2 for a lower energy cost than in SiO_2 [9].

It is possible that the key defects in SiO_2 and GeO_2 are different. In SiO_2, the main defect is the O vacancy or E' center, and related defects such as the single Si dangling bond. There is also the Si dangling bond on the Si side of the Si/SiO_2 interface (P_b center). In GeO_2, oxygen deficiency is a critical parameter. The Ge^{2+} state is more stable in GeO_{2-x} than the equivalent Si^{2+} in SiO_2, for example the Ge^{2+} state is employed in GeO_2 based Bragg gratings [9]. The Ge^{2+} state is also notable in the greater volatility of molecular GeO, as isotope tracer studies confirmed the O lattice diffusion through the GeO_2 layer [10]. We note that the Ge^{II} oxidation state is widely seen in Ge chemistry, such as in GeSbTe phase change memory materials [11].

In terms of reliability, Lu et al [12,13] recently noted that the addition of group IIIA oxides such as Y_2O_3 into GeO_2 can be beneficial. This reduces the GeO_2 etch rate and GeO evolution rate from GeO_2. The trap density D_{it} and the CV hysteresis are reduced and the electrical reliability is improved. Since the interfacial defects usually act as interfacial traps and cause hysteresis in CV measurer, the improvement in CV and D_{it} was attributed to the doping effect on defect control. This effect was explained empirically in terms of a more rigid oxide network which led to a lower defect generation rate [13]. However, in gate oxides, reliability is generally related to charge trapping at *existing* defects, not the generation of new defects [6,14]. Thus, an atomistic model of the effects would be useful.

Here, we study the electronic structure and energies of the O vacancy and Ge^{2+} defects in GeO_2 compared to those in SiO_2 using density functional theory (DFT) calculations. We explain the effect of Y additions as well as other rare earth element doping in GeO_2, and find that it leads to a passivation effect, expelling defect states out of the GeO_2 band gap. This suggests that the improved reliability of Y doped GeO_2 arises from the passivation of charge traps in GeO_2. The overall effect is compared to that of La, Y or N dopants on O vacancy behavior in HfO_2 where

defect states are expelled from the gap so that the electronic structure reverts to the more stable closed-shell configuration.

2. Calculation setup

Our simulations are carried out using the plane-wave, density functional code CASTEP [15], using norm-conserving pseudopotentials with a plane-wave cutoff energy of 750 eV. A Γ point averaging scheme is chosen in all calculations due to the large supercells used. The geometry optimization task is carried out using the generalized gradient approximation (GGA) for the electronic exchange-correlation functional until the residual force on each atom is below 0.01eV/Å. The GGA is known to under-estimate the band gap, so the electronic density of states is calculated using the screened exchange hybrid functional [16] to correct the band gap error of GGA.

An amorphous GeO_2 model of 144 atoms was generated by several separate steps. First, a random network of GeO_2 is created by a high temperature molecular dynamics (MD) anneal of crystalline (quartz-like) GeO_2 at 2000 K for 20 ps. It is then cooled down to 300 K in 20 ps. Finally, the system is relaxed to an energy minimum.

We then generated a smaller 72 atom supercell of GeO_2 for the alloying studies with similar steps. To form the alloy models, we replace pairs of Ge's with trivalent Y, La, Sc, Lu or Al atoms. It is necessary to also remove one oxygen per pair of metal atoms to keep valence balance. An important point is that the coordination numbers of the Sc, Y and La metal atoms are much larger than those of Ge, so a good alloy structure requires careful relaxation so that the system is not trapped in a metastable minimum of a substituted network.

3.Results and Discussion

3.1 Amorphous models of GeO_2 and doped GeO_2

Fig. 3.1 shows the supercell of the random network of GeO_2. All the Ge sites are 4-fold coordinated and all the O sites are 2-fold coordinated, just as in an a-SiO_2 network. **Fig 3.2(a-c)** show the networks containing 8%, 16% and 33% Y loading respectively, as a fraction of the cation sites. It is seen that the Y sites tend to be paired close by each other. The coordination numbers of the metal sites are given in **Table 3.1**, being 6.25 for Y sites. This is achieved by the Y atoms making additional bonds to oxygens, converting those oxygens into 3-fold sites.

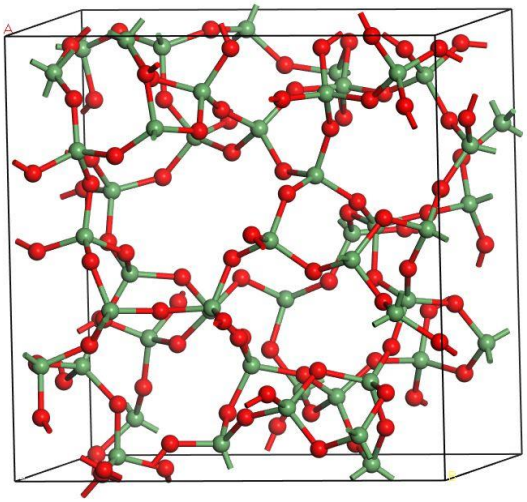


Fig 3.1 Random network model of amorphous GeO_2. (green balls = Ge, red balls = O)

(a) (b) (c)

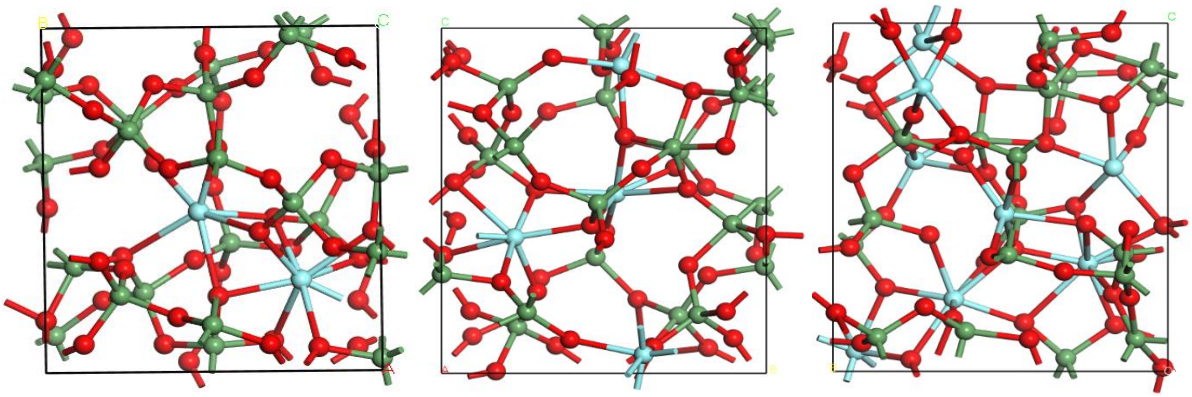


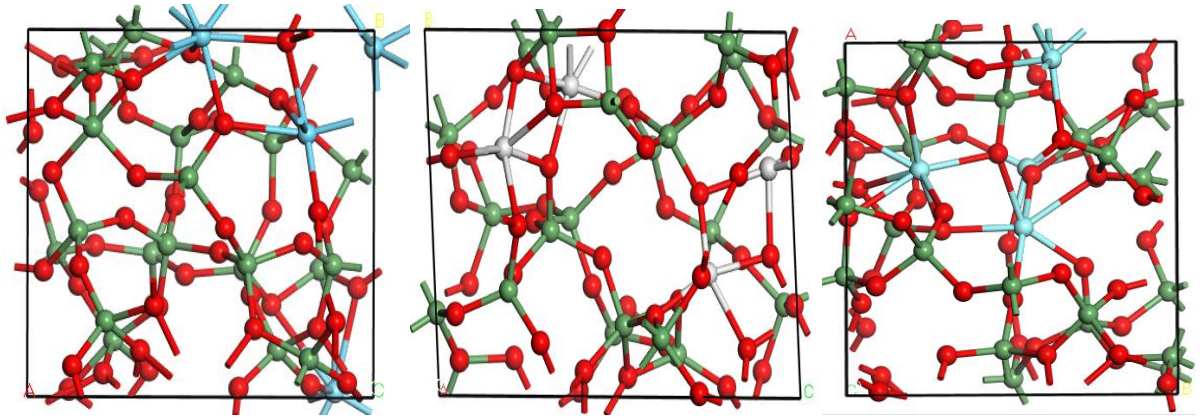
Fig 3.2 network model of GeY_xO_2 with (a) 8% Y, (b) 16% Y, (c) 33% Y. (green balls = Ge, red balls = O, blue balls = metal)

Table 3.1 Metal site coordination number in GeO_2.

Dope metal	Sc	Y	La
Coordination number	5.25	6.25	6.75

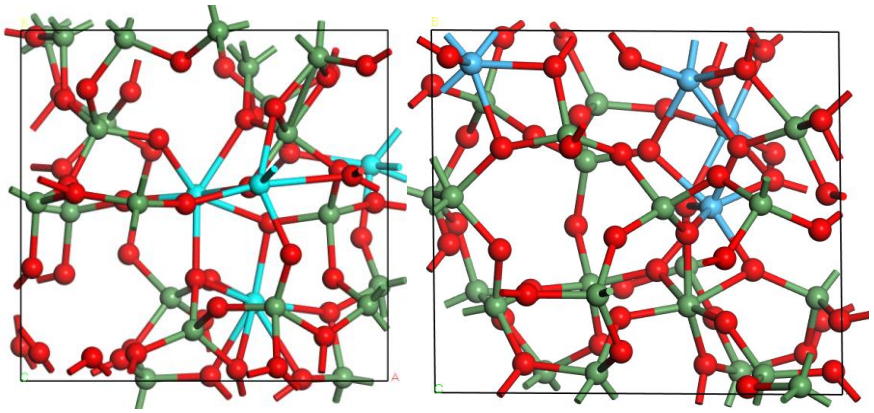
The metals Sc, Y and La are highly electropositive atoms with large ionic radii. **Fig 3.3** shows the network structure of the 16% metal alloy for the three different trivalent metals. As the metal changes from Sc to Y to La, we see that the metal-O coordination increases with atomic number and ionic radius, as summarized in **Table 3.1**. Thus, there is some dependence of metal coordination on ionic radius. HfO_2 alloying is also included, in that case Hf is tetra-valent like Ge, so there is no need to remove an oxygen for valence balance.

(a) (b) (c)



Sc doped Y doped La doped

(d) (e)



Lu doped Hf doped

Fig 3.3 Random network models of $GeMe_xO_2$ with 16% metal, for Sc, Y, La, Lu and Hf doping, respectively. (green balls = Ge, red balls = O, blue balls = metal)

Fig 3.4 plots the atomic volume per oxygen atom of the Y alloy as a function of the cation fraction. It is seen that Y addition causes a decrease in mean atomic volume below the linear interpolation line of Vegard's law[17], corresponding to a downward bowing. This is as expected for exothermic mixing. The exothermic mixing of Y_2O_3 and GeO_2 means that Y_2O_3 behaves like an oxide base towards GeO_2 [18-20]. This is because of the large Y ionic radius.

Lu and Toriumi [21] proposed that various features of electrical reliability might arise from an increase in rigidity of the bonding network, which reduces the rate of defect formation. We therefore calculated the bulk modulus of the oxide alloys as a function of metal content. **Fig 3.5** shows that the bulk modulus increases progressively with Y addition up to 50% metal. We

also calculated the bulk modulus of crystalline (hexagonal) Y_2O_3 and pyrochlore $Y_2Ge_2O_7$, for reference. We see that the amorphous phases have much lower bulk moduli than the crystalline phase of the same composition.

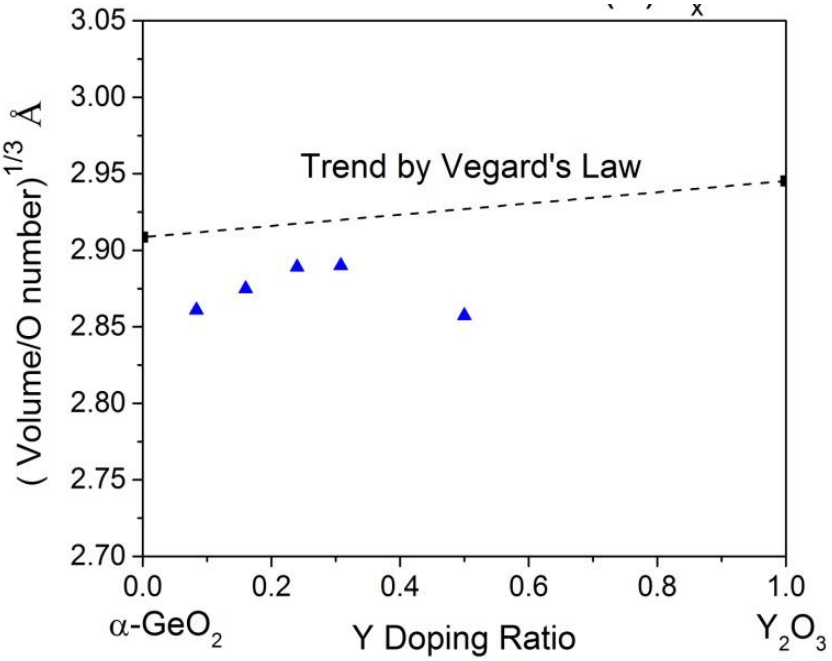


Fig 3.4 Calculated atomic volume per O atom, vs Y doping ratio.

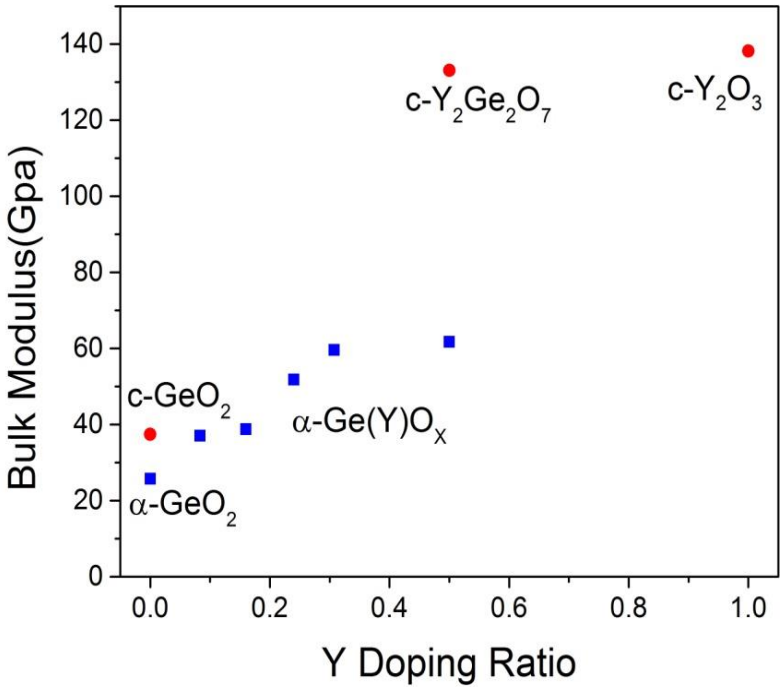
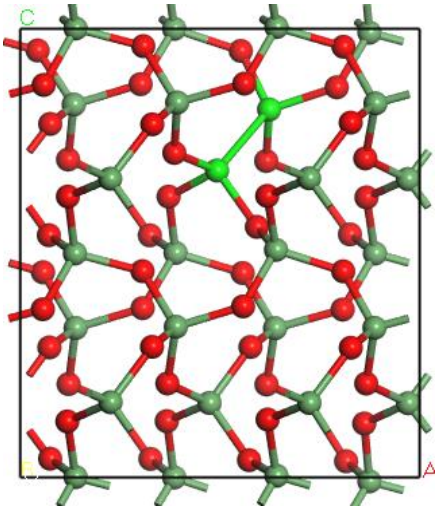


Fig 3.5 calculated bulk modulus of a-GeY_xO_2 compared to modulus of crystalline GeO_2, Y_2O_3 and $Y_2Ge_2O_7$.

3.2 Oxygen vacancy in undoped and doped GeO_2

We now consider the oxygen vacancies. **Fig 3.6(a)** shows that the neutral O vacancy in pure GeO_2 relaxes and reconstructs to form a Ge-Ge bond. This behavior is like the O vacancy in SiO_2 [28]. The Ge-Ge bond gives rise to gap states, as seen in the sX partial density of states in **Fig 3.6(b)**, with a filled bonding state at 0.8 eV above the bulk valence band maximum (VBM) at 0 eV and an empty anti-bonding state at +6.8 eV, that is just above the conduction band minimum given a GeO_2 band gap of 6.2 eV. The formation energy of the relaxed O vacancy in a-GeO_2 is about 3.33 eV, **Table 3.2**. Clearly this energy forms a distribution, due to disorder, but its energy range is not so large, as the coordination does not vary.

(a)



(b)

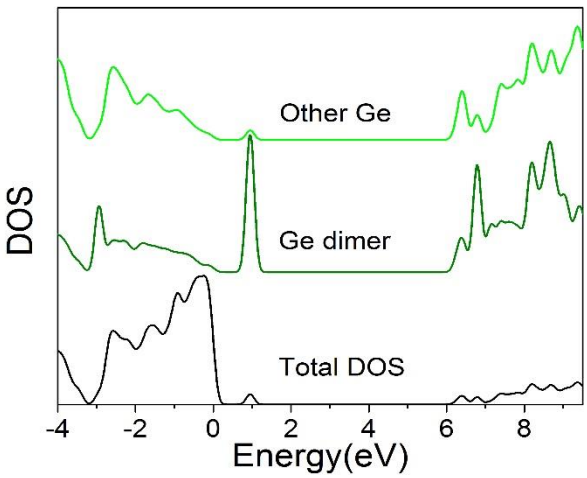
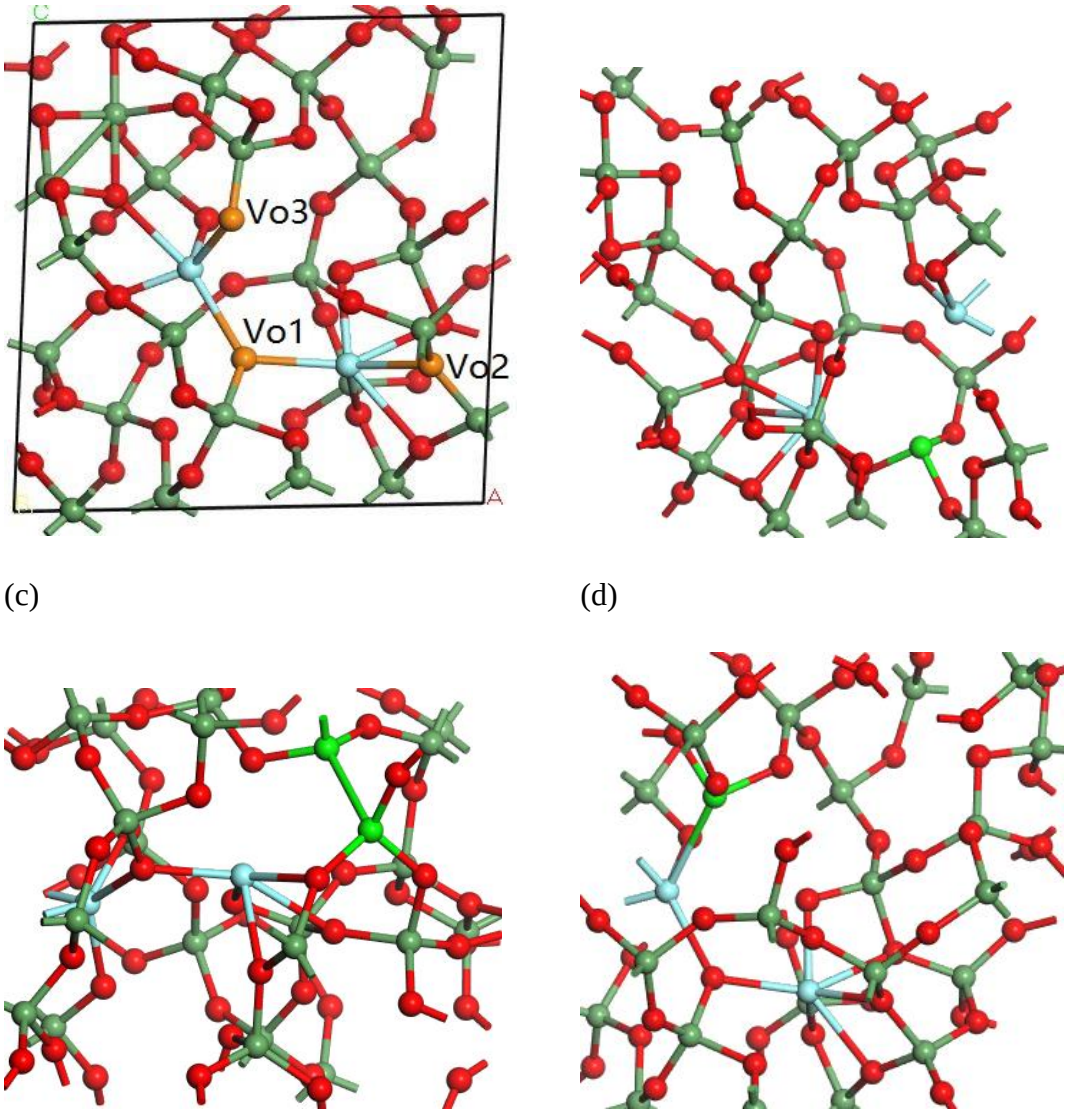


Fig 3.6 (a)GeO_2 network, containing relaxed O vacancy (Ge-Ge bond). (b) Calculated sX partial density of states (PDOS) of a Ge-Ge bond ion a-GeO_2.

We then calculated the formation energy for the neutral O vacancy in the GeY_xO_y alloys for different compositions. There are three different O vacancy sites in the GeY_xO_y alloys as shown in **Fig 3.7(a)**, (i) an O bonded to two Y and one Ge, (ii) an O bonded to one Y and two Ge's, and (iii) an O bonded to one Y and one Ge. The fraction of type (iii) sites declines to 0% at about 33% Y content.

(a) (b)



(c) (d)

Fig 3.7 (a) the three O sites in GeY_xO_2. (b-d) The relaxed geometries of the O vacancies at these 3 sites. Illustration of an O site that is a second neighbor to a O vacancy next to an Y atom. (green balls = Ge, red balls = O, blue balls = metal)

Table 3.2. Formation energy (eV) of the O vacancy in GeO_2 and the 3 types of oxygen vacancy in GeY_xO_2, in the O-rich limit, and of a bulk O vacancy that is a second neighbor of a Y site.

	Pure GeO_2	Vo1	Vo2	Vo3	V(2nn)
Formation energy	3.33	4.27	4.67	3.04	3.44
Diffusion energy	5.70	6.53	5.97	5.65/6.10	

Fig 3.7(b-d) show the relaxed geometries of these three vacancy sites. We see that vacancies (i) and (iii) relax their bond angles, but there is no reconstruction or rebonding across the vacancy for Y. In contrast, for vacancy (ii), the two Ge dangling bonds rebond into a Ge-Ge bond, just as in the case of the O vacancy in GeO_2 itself.

Table 3.2 shows the defect formation energies for these relaxed vacancies next to Y sites, in the 8% Y network, which are 4.27 eV, 4.67 eV and 3.04 eV for vacancies Vo1, Vo2 and Vo3, respectively. This is roughly in line with the number of bonds broken to create the vacancy. Note that the largest formation energy is for vacancy Vo2 with two adjacent Ge sites, despite 'getting energy back' for reforming the Ge-Ge bond. The calculation is for an O chemical potential of 0 eV, that is assuming the removed O atoms form O_2 molecules. Thus, on average, there is a substantial 0.66 eV average increase in the O vacancy formation energy in GeO_2 by adding the Y. This average is also consistent with the average O vacancy formation that we calculate for O sites in the pyrochlore $Y_2Ge_2O_7$ crystal. The O vacancy formation energy in $Y_2Ge_2O_7$ at its three different sites is 4.29 eV, 4.92 eV and 4.29 eV, or 4.50 eV on average. The average formation energy increases with increased alloying, because the fraction of Vo3 sites goes down, and Vo2 then domintaes.

We also calculated the diffusion energy for each type of O vacancy. This is the sum of formation energy and migration barrier. The migration barrier is calculated ranging from the O vacancy in pure GeO_2 (which is relatively constant) to each of the O vacancy states in GeYOx. We see that the diffusion energy increases by 0.45eV.

The increases in vacancy formation energy and migration energy at O sites adjacent to Y sites partly explains the reduced molecular GeO evolution rate in the alloys. However, the data shows that the rise in evolution temperature occurs already in alloys with only 10% Y [22]. This Y addition would only affect vacancies at 30% of oxygen sites (assuming 3-coordinated

oxygens). A percolation path for O diffusion along O sites would not be blocked as it is below the roughly 55% percolation threshold for this lattice.

Table 3.3 Number of different vacancy sites in GeYOx vs Y content, in 72 atom supercell. Type Vo1, Vo2, Vo3 are shown in **Fig 3.7**. Vo4 is an oxygen site with 4 or 5 neighbors, found in Y-rich alloy networks.

	8%, 2Y	16%, 4Y	25%, 6Y	33%, 8Y
Vo1	3	5	8	11
Vo2	6	9	15	12
Vo3	2	6	6	4
Vo4	0	0	0	6
Total	11	20	29	33

Thus, we also calculated the vacancy formation energy at O sites that are second nearest neighbors to Y sites, **Fig 3.8**. We found the O vacancy formation energy increased from 3.33 eV for pure GeO_2 to 3.44 eV for the second neighbor sites, a moderate increase. An effect at the second neighbor distance increase the effect's range by a factor of roughly three over the nearest neighbor effect, of the Y-O bond length, or roughly 9 times in terms of projected area, as illustrated in **Fig 3.8**. Now the addition of 10% Y will have a full blocking effect on O vacancy diffusion, by percolation, consistent with experimental observations.

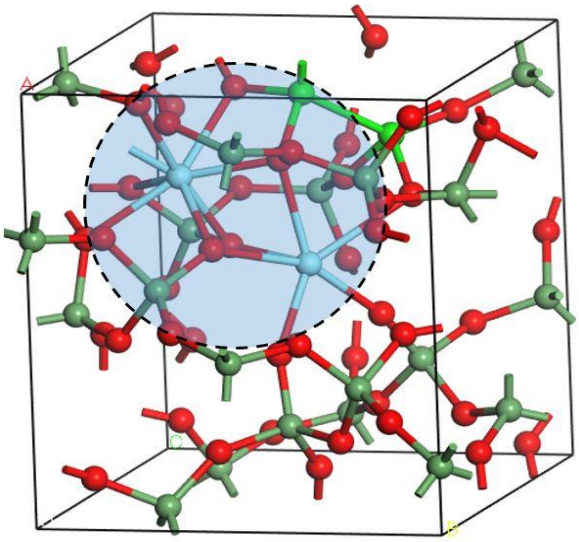


Fig 3.8 The relaxed geometries of the O vacancies at the second neighbor site, transparent blue area shows the region influenced by this second neighbor effect.

We now consider the electronic structure of the vacancies. **Fig 3.9(a-c)** shows the calculated partial density of states on atoms at the three different vacancy sites, Vo1, Vo2, Vo3. For Vo1, the single Ge dangling bond (DB) gives rise to a filled state for a negative DB lying at -0.2 eV just below the top of the valence band. On the other hand, the Y sites adjacent to the vacancy give rise to no gap states, their PDOS starts at 3.3 eV and above. In effect, they form positive Y dangling bond states in the conduction band. For Vo2, the Ge-Ge re-bond gives rise to a filled bonding state in the valence band, and an antibonding state at +1.7 eV in the Ge conduction band, and no gap states. In Vo3, again the Ge DB site next to the vacancy gives rise to a filled defect state at 0 eV, at the top of the valence band.

Table 3.4 Formation energy (eV) of the V1 oxygen vacancy in $GeMe_xO_2$, in O-rich limit, as a function of the alloying metal (Me) oxide.

Sc	Y	La	Lu	Hf
3.47	3.77	3.61	3.74	4.69

We also calculated the O vacancy formation energy as a function of metal. Limiting ourselves to the vacancy V_{O1}, we see that its structure depends on the metal, as shown in **Fig 3.10**. The formation energy increases from Sc to Y, but it then decreases for La. This is because the large ionic radius of La allows an extra La-O bond to form across the vacancy (circled), and this decreases the formation energy (**Table 3.4**). The Ge dangling bond atoms are shown in bright green in these figures.

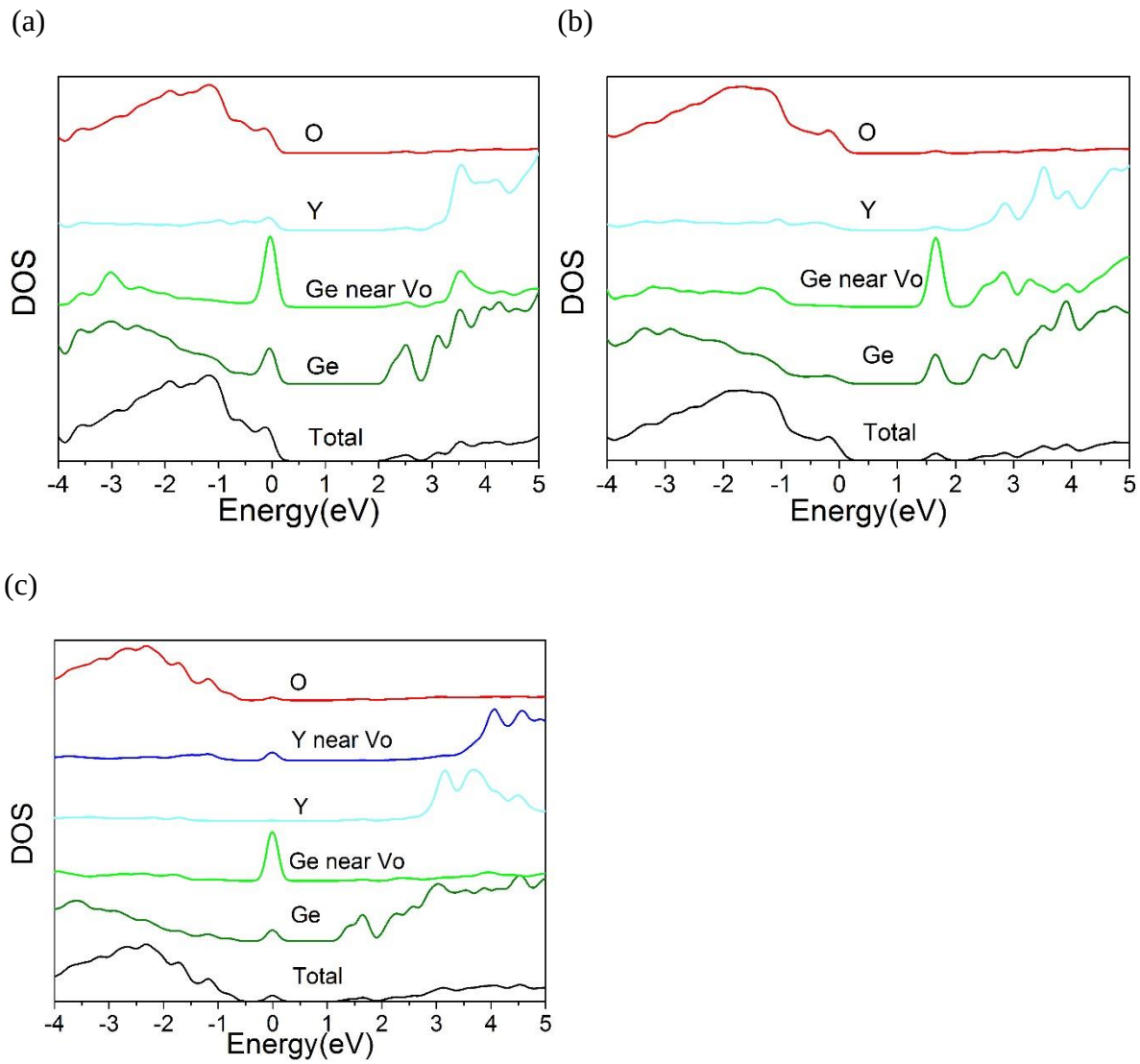
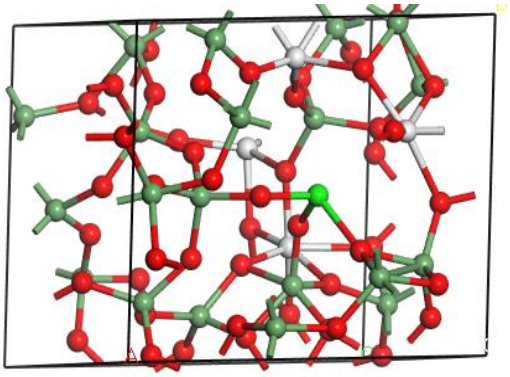
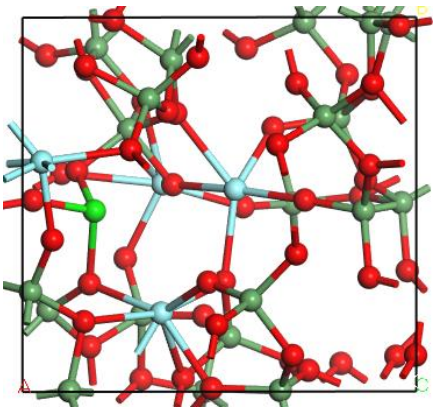


Fig 3.9 Calculated sX partial density of states (PDOS) of atoms near the three O vacancies (a) Vo1, (b) Vo2, (c) Vo3, as defined in **Fig** 3.7.

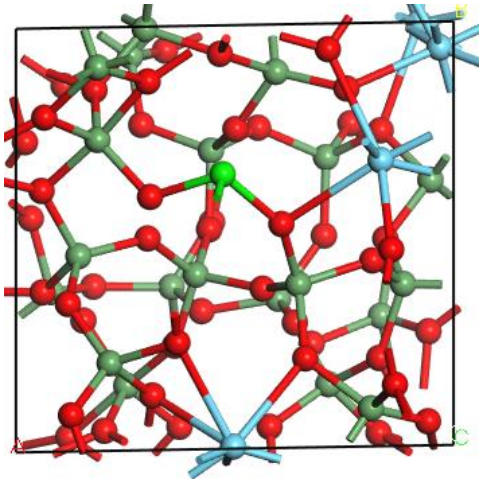
(a)



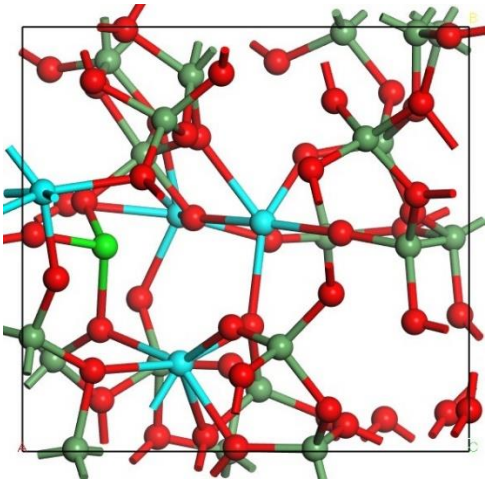
(b)



(c)



(d)



(e)

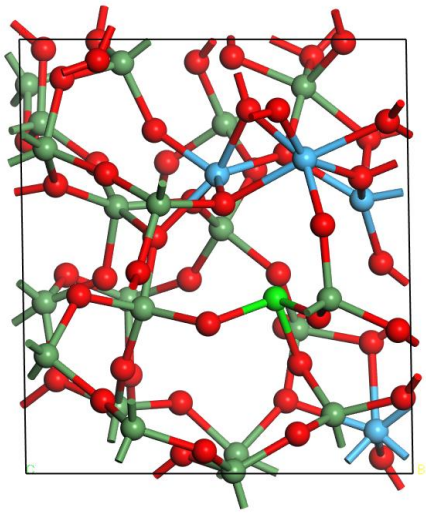


Fig 3.10 Oxygen vacancies in GeO_2 alloyed with Sc, Y, La, Lu, and Hf oxide. Any Ge dangling bond site next to vacancy is shown as bright green.

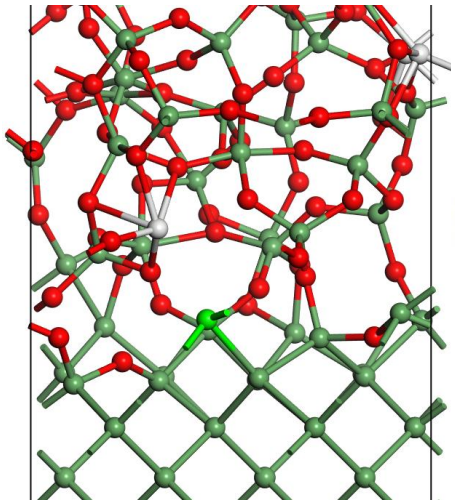
3.3 Oxygen vacancies at undoped and doped Ge:GeO_2 interface

We now consider some possible disadvantages of metal addition, by calculating the effects on O vacancies at the Ge/GeO_2 interface. We saw from **Fig 3.6** that an O vacancy in bulk GeO_2 does not form gap states in its simple configuration. The main question is whether O vacancies between a metal atom and the Ge side will form gap states. Generally this will occur if a metal-Ge bond forms [23-25].

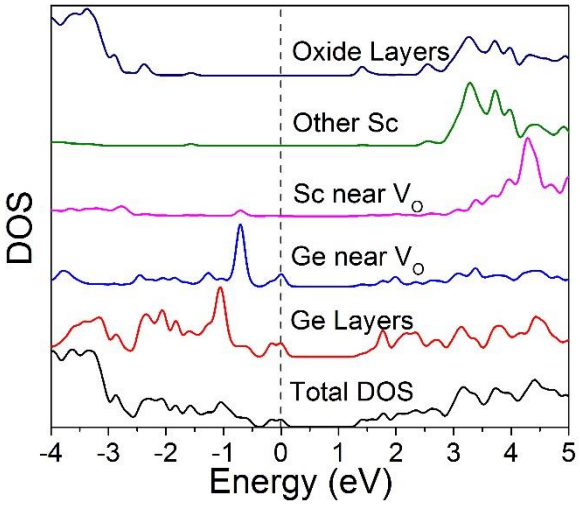
We construct a Ge/GeO_2 interface model with 48 Ge atoms in 6 layers, a (100) interface, 36 formula units of amorphous GeO_2, and terminated on both top and bottom by hydrogens, with 15Å of vacuum. The Ge sites on the Ge side of the interface each have two dangling bonds pointed towards the GeO_2 side. One of these at each site forms a Ge-O-Ge bridge with an adjacent Ge parallel to the interface, leaving the remaining DB to form a Ge-O bond across to the GeO_2, as in the Yu and Tersoff [26] model. This is the basic interface structure. We then replace a Ge site within one O of the interface by a metal atom. An additional H atom is added to the GeO_2 side if the metal atom is trivalent, for valence balance.

An oxygen atom between the metal atom and the Ge interface atom is then removed to make the vacancy, and the structure allowed to relax. **Fig 3.11** shows the structures of the relaxed interfacial O vacancy for the different $MGeO_x$/Ge interfaces. We see that a metal-Ge bond does not form in the cases of Sc, Y and Al, but it does form in the cases of La and Hf. This is slightly different to that found previously by Dimoulas [25]. For Hf, it is similar to that found in previously [27]. There is a clear gap state at +1 eV for the Hf case. These results can be summarized in a plot of ionic radius vs interfacial defects, **Fig 3.12**. Lu et al. suggested two criteria for the choice of doping metal [13]. First, the doping metal atom should be large enough to influence the GeO_2 network and increase the stability. Moreover, the doping atom should not be too reactive with Ge in experiment that deteriorate the electrical properties of GeO_2. However, it is unclear how the doped metal elements deteriorate the electrical properties. In our simulation, we confirm that the reactivity of Ge with the metal corresponds to its ability to form a Ge-metal bond across the vacancy, and such defect is responsible for the poor electrical properties of the doped GeO_2 found in Lu's experiment[13]. It is interesting that there is subtle dependence of this energy on the ionic radius of the group IIIA metal, so that Sc or Y are favored by not forming a bond, while La is too large and reactive.

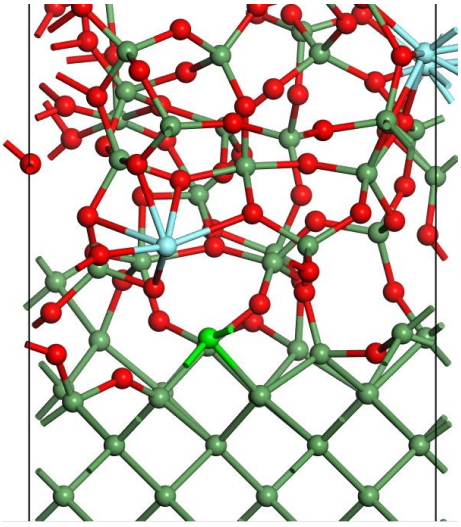
(a)



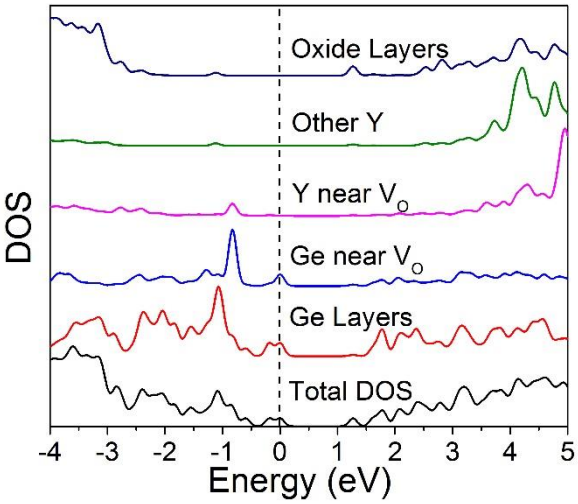
(b)


Oxide Layers
Other Sc
Sc near V_O
Ge near V_O
Ge Layers
Total DOS
DOS
-4 -3 -2 -1 0 1 2 3 4 5
Energy (eV)

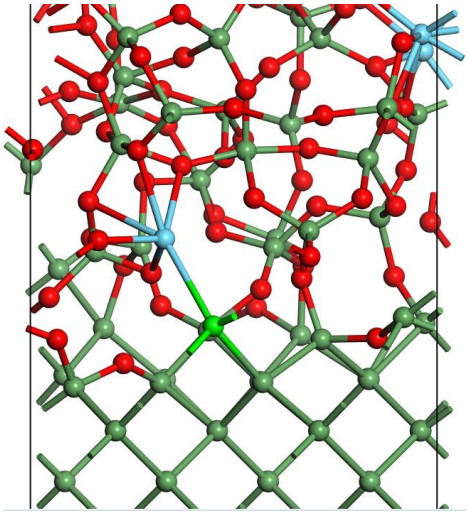
(c)



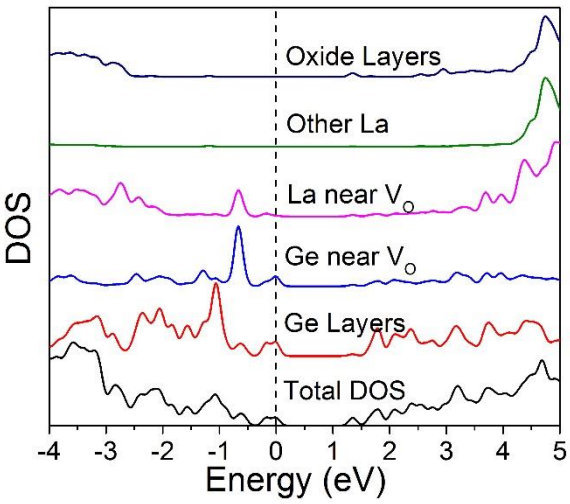
(d)


Oxide Layers
Other Y
Y near V_O
Ge near V_O
Ge Layers
Total DOS
DOS
-4 -3 -2 -1 0 1 2 3 4 5
Energy (eV)

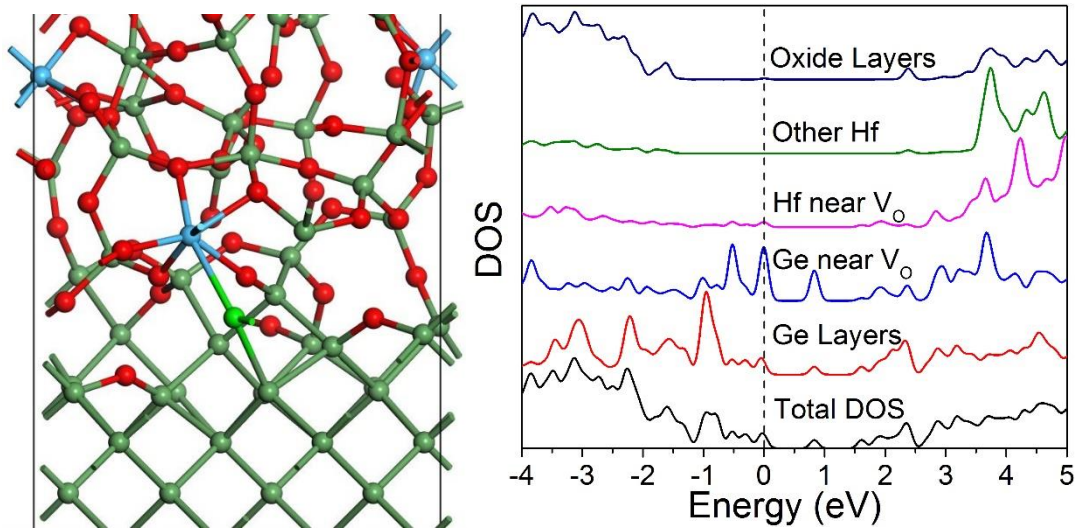
(e)



(f)


Oxide Layers
Other La
La near V_O
Ge near V_O
Ge Layers
Total DOS
DOS
-4 -3 -2 -1 0 1 2 3 4 5
Energy (eV)

(g) (h)



(i) (j)

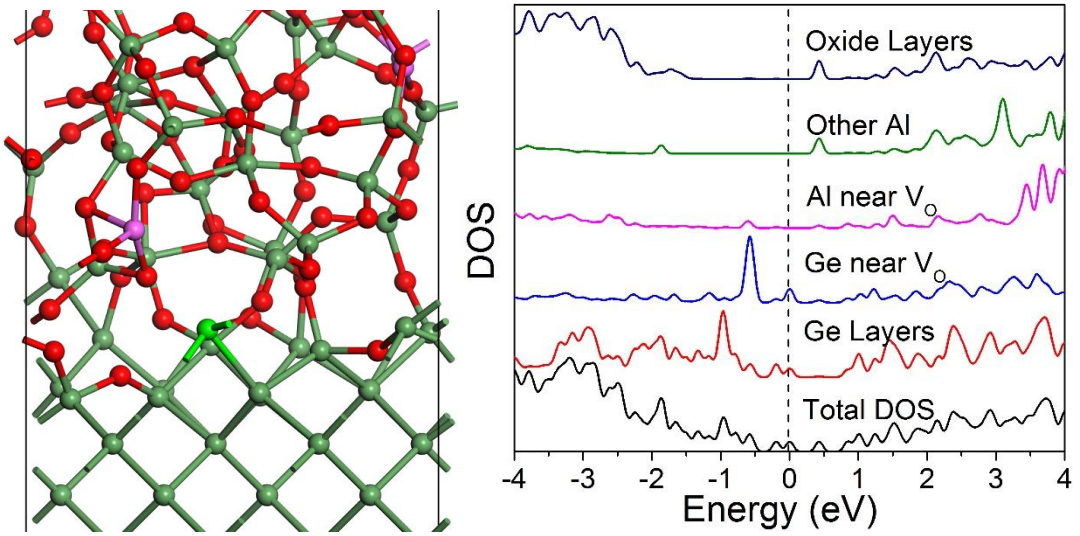


Fig 3.11. Geometries of interfacial O vacancies and local PDOS of adjacent sites for Sc, Y, La, Hf, an Al oxide alloys on Ge. Note no bond reforming for Sc, Y, or Al. Ge dangling bond atoms are in bright green color.

Table 3.5 Formation energy of interstitial O vacancy for different metal additions.

Doped metal	Sc	Y	La	Al	Hf
Formation energy	2.95	2.55	2.51	3.45	3.37

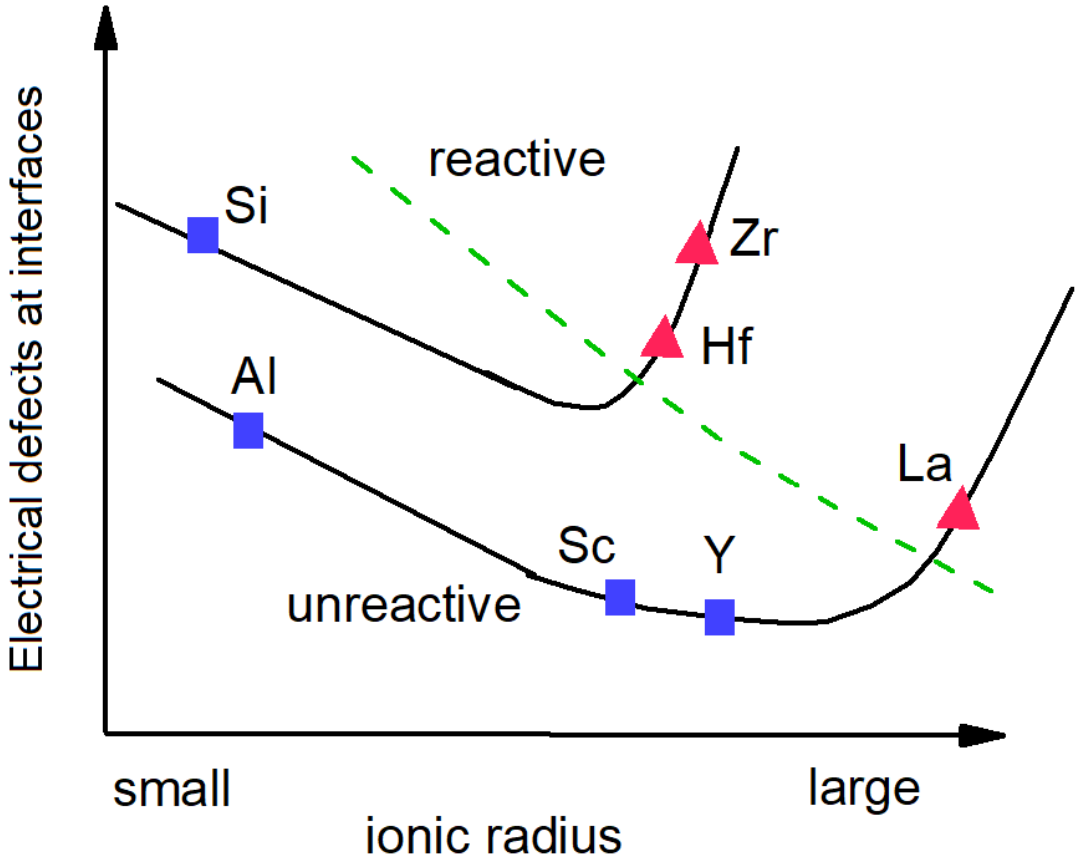


Fig 3.12 Schematic of Lu/Toriumi diagram to show reative/unreactive interfaces for metal oxide/GeO_2 alloys on Ge, as defined against (experimental) electrical properties and ionic radii. Figure adapted from Ref[13]

3.4 Valence alternation pair O deficiency defect

(a) (b)

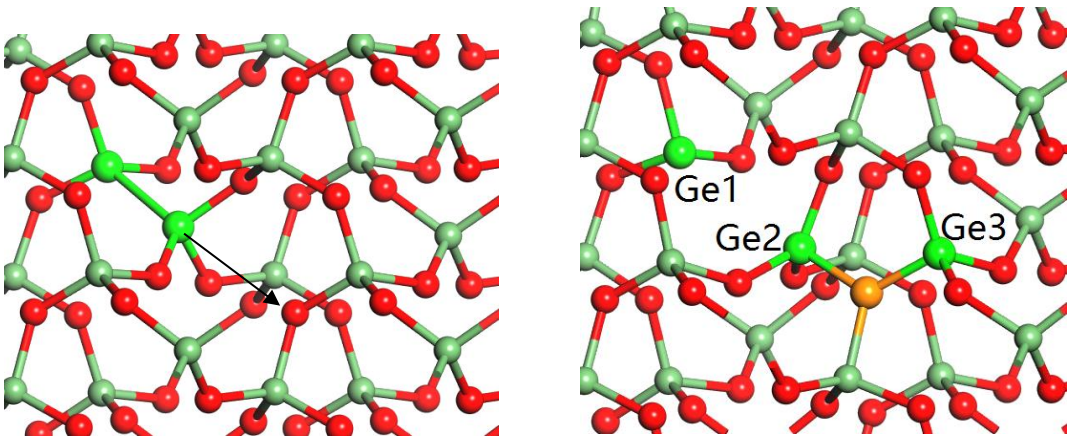


Fig 3.13 (a) The reconstructed O vacancy (in the form of Ge-Ge dimer, shown with bright green color) in crystalline GeO_2, and (b) the valence alternation pair or VAP, made by inversing

the Ge2 site (the bright green atom in the middle) and its bonding to the O. Green = Ge, bright green = Ge atoms involved in defects, red = O, and orange = O at VAP.

We first consider the case of the neutral O vacancy in crystalline (c-) GeO_2. Creating the vacancy leaves two Ge dangling bonds, which reconstruct to form a Ge-Ge bond, as in SiO_2 [28], **Fig 3.13(a).** The formation energy of the vacancy in oxygen-rich conditions (atmospheric pressure O_2) is given in **Table 3.6**. We see that the vacancy formation energy is much greater in SiO_2 than in GeO_2. Its formation energy in the O-poor condition (the Ge/GeO_2 or Si/SiO_2 equilibrium) is given by subtracting the bulk heat of formation of GeO_2 or SiO_2 per O atom, and is much lower.

Table 3.6. Summary of O deficiency defect formation energy (in eV) in bulk SiO_2 crystal, GeO_2 crystal, amorphous GeO_2, and at the Ge/GeO_2 interface in the O rich condition. Formation energies in the O-poor condition are found by subtracting the heat of formation of GeO_2 or SiO_2 per oxygen.

Oxygen deficiency defect type		c-SiO_2	c-GeO_2	a-GeO_2	Ge/a-GeO_2 interface
O vacancy	O-rich	6.14	3.16	3.04	3.05
	O-poor	1.29	0.14	0.13	0.14
VAP	O-rich	9.34	5.01	4.76	3.45
	O-poor	4.49	1.99	1.85	0.54

The Ge^{2+} configuration can be given by a variant of the valence alternation pair (VAP) defects found in chalcogenides. In a covalent random network of a-Se, a VAP consists of an under-coordinated and an over-coordinated atom [29]. A Se-Se bond is broken, and one of the resulting 1-fold coordinated Se sites rebonds to a bulk 2-fold Se site to make it 3-fold site. To do this, the site must become positively charged, then it can form 3 bonds by the 8-N rule. The VAP defect consists of Se_3^+ and Se_1^-, where the subscript denotes coordination number. The VAP is considered to be a low energy defect because the total number of bonds is conserved [17]. VAPs can also exist in binary networks, and have been suggested for SiO_2 [28,30-32] although the vacancy model is generally favored. The VAP is one way to represent the O deficiency of divalent Ge, as explained below.

Starting from the reconstructed O vacancy with a Ge-Ge bond as in **Fig 3.13(a)**, we break the Ge-Ge bond and flip one of the 3-fold Ge atoms away from the vacancy towards an oxygen atom on its other side to form a new Ge-O bond. This process is shown in **Fig 3.13(b)** for

crystalline GeO_2 where it is easier to follow, and in **Fig 3.14(a)**, **Fig 3.15(a)** for a random network. For this defect to be consistent with the 8-N rule, the 3-fold O site must become positive as in O_3^+. Its lost electron transfers to the other 3-fold Ge site to make it Ge_3^-. These defects have been studied by Binder et al [33,34] and Li and Lin [27,35,36]. Here, the total number of bonds is conserved, but a Ge-Ge bond is replaced by a Ge-O bond.

(a) (b)

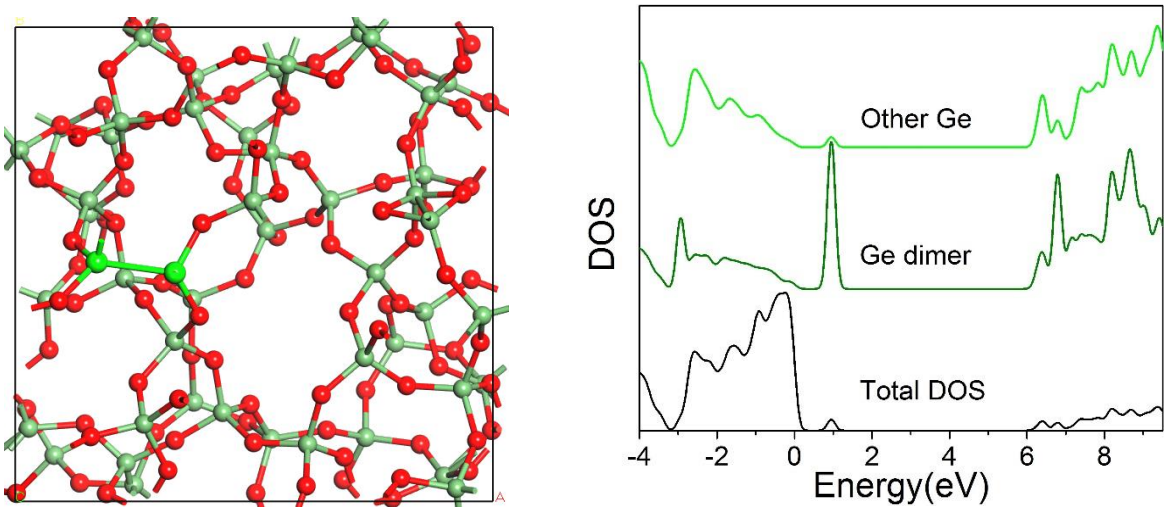


Fig 3.14 (a) Amorphous GeO_2 with an oxygen vacancy reconstructed as a Ge-Ge bond; (b) the related PDOS.

(a) (b)

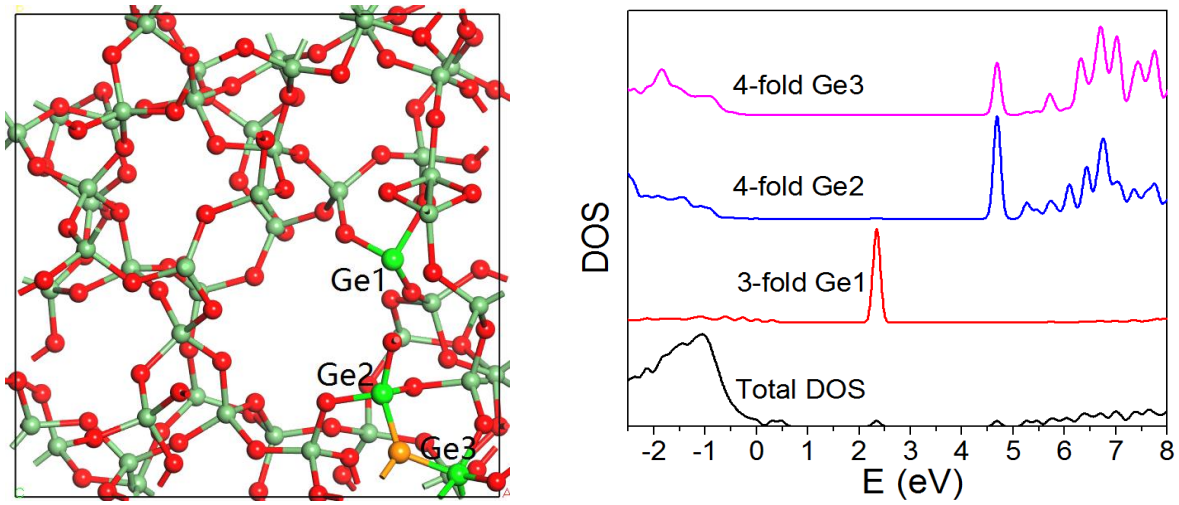


Fig 3.15 (a) Amorphous GeO_2 with a VAP type oxygen deficient defect; (b) the related PDOS.

The formation energy of the VAP in SiO_2 and GeO_2 are calculated using supercell models and these are given in **Table 3.6**. We see that the VAP is ~1.8 eV more expensive than the O

vacancy in a-GeO_2, but it is ~3.2 eV more expensive in a-SiO_2. The cost of the vacancy and the VAP become much lower in the O-poor condition, as occurs at for example the Ge/GeO_2 interface. There are also a range of formation energies for defects at the Ge/GeO_2 interface because of the wide range of possible bonding configurations there.

The local partial density of states (PDOS) has been calculated for these defects, using the sX functional. **Fig 3.14(b)** shows the PDOS for the Ge-Ge bond across the reconstructed O vacancy in a-GeO_2. It gives a filled bonding state at 1 eV above the GeO_2 VBM, and an empty antibonding state at 6.2 eV in the GeO_2 conduction band. These states are aligned with the band edges of the underlying Ge layer in **Fig 3.19** using the conduction band offset of 0.9 eV [27]. We see that these defect states are not so close energetically to the Ge gap, so they are less critical for charge trapping.

(a) (b)

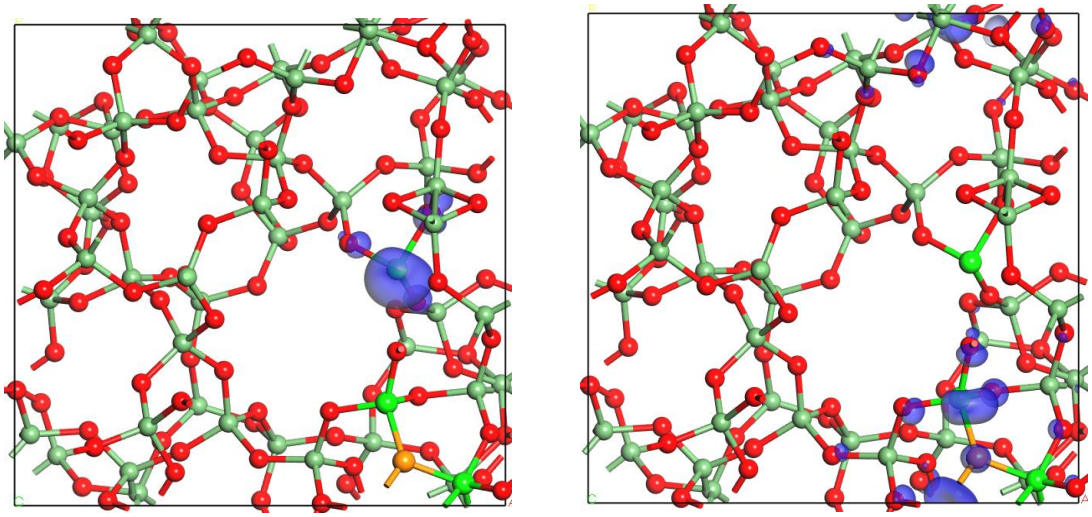


Fig 3.16 (a) Occupied defect state at 2.6eV, and (b) empty defect orbital at 5.5eV of VAP in amorphous GeO_2.

In the VAP, the 3-fold Ge^- atom is labelled Ge1 in **Fig 3.13(b)** and the other Ge atom previously at the vacancy is labelled Ge2. The O_3^+ site lies between the sites labelled Ge2 and Ge3. The PDOS of the VAP has two gap states, see **Fig 3.15(b)**. The filled state at 2.6 eV in midgap consists of a doubly occupied dangling bond orbital on the Ge_3^- (Ge1) site. Its wavefunction in **Fig 3.16(a)** confirms that it is strongly localized on the Ge1 atom. The second empty gap state at 5.5eV lies just below the GeO_2 CBM. This gap state formally derives from the 3-fold O atom. However, because of the polarity of the Ge-O bonds, this orbital is mostly localised on two adjacent Ge atoms Ge2 and Ge3, see **Fig 3.16(b)**. These two gap states lie much closer to the

band edge energies of Ge itself than those of the O vacancy, as summarised in **Fig 3.19**. Thus, these defects act as more effective charge traps, and contribute towards D_{it}. If the VAP defects are located at the interface, they can act as border traps, typically with a broader energy distribution which tails into the Ge gap.

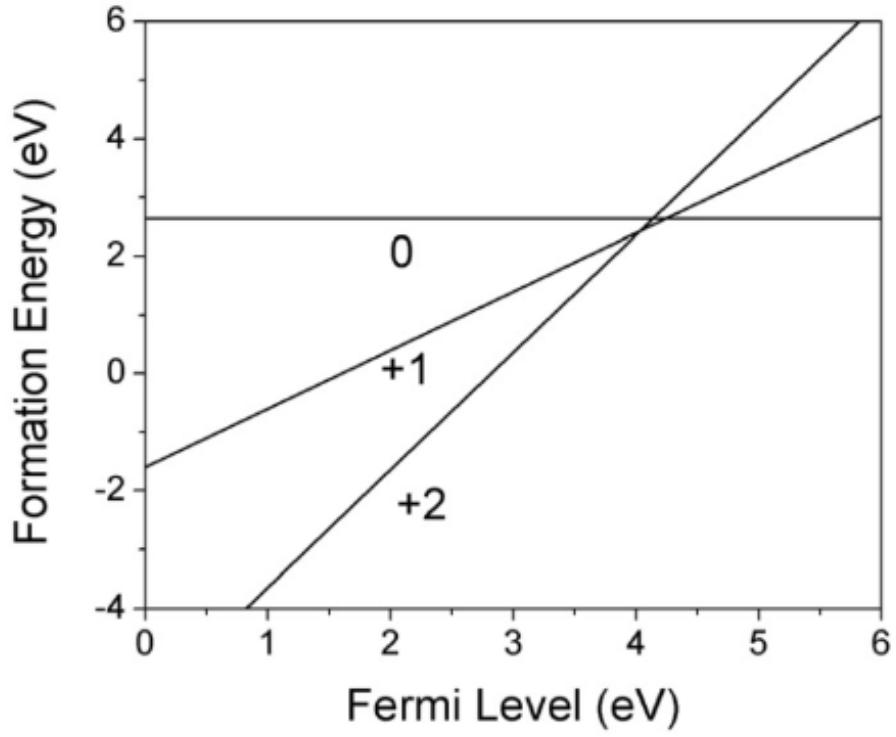


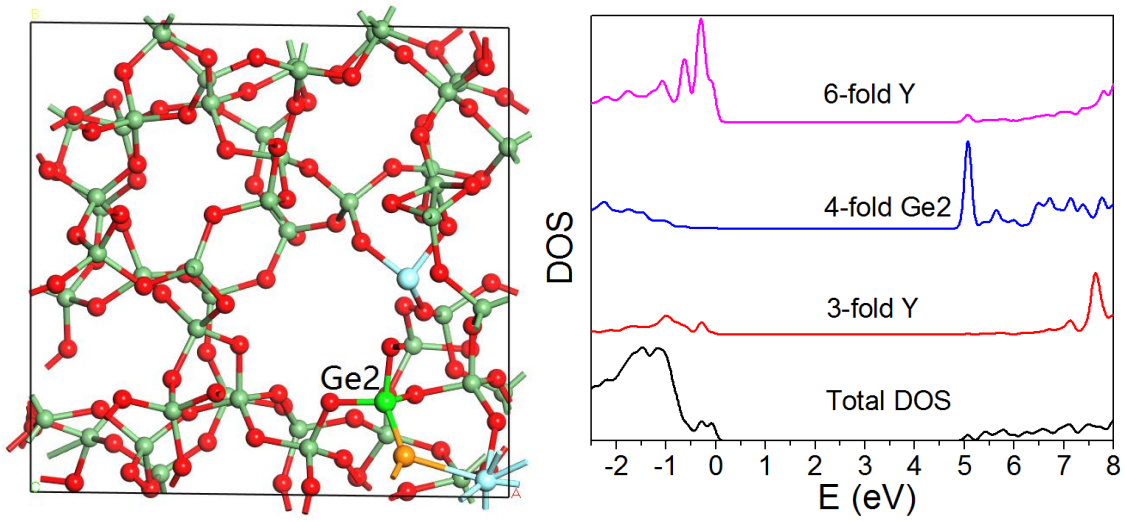
Fig 3.17 The relative defect formation energy vs Fermi Level for the VAP. To clarify, the charge on Ge_1^+, or Ge_3^-, is the chemical formal charge, whereas the 0, +1, +2 charge as shown in this figure are the overall defect charge state of the VAP.

3.5 Passivation of VAP in amorphous GeO_2 by Y doping

We now consider a GeO_2 alloy in which two Y atoms are substituted for the two Ge atoms involved in the VAP, **Fig 3.18(a)**. We replace the Ge1 and Ge3 atoms by Y, while the Ge2 atom remains unchanged initially to keep the Y's separate. The network is relaxed. After geometry optimization, the Y replacing 3-fold Ge1 site remains 3-fold, while the Y replacing the 4-fold Ge3 becomes 6-fold coordinated by the formation of extra Y-O bonds to neighbouring oxygens. The PDOS of these sites in **Fig 3.18(b)** shows that the filled defect state at 2.6 eV on Ge1 is now removed from the gap, while the defect state at 5.5 eV now moves across to the Ge2 site. Overall, the Y substitution has passivated the hole trap at 2.6 eV by removing this state from the gap into the conduction band, as summarised in **Fig 3.19**. Thus, there is no longer a hole trap to lead to the poor reliability.

Now we consider the defect passivation mechanism. The defects give a gap state filled with 2 electrons. Substituting two Ge's with Y's contributes two fewer electrons, that is it forms two holes in the valance band. Removing that gap state allows the two defect electrons to fall into the 2 holes in the valence band, regaining a more stable closed-shell configuration.

(a) (b)



(c)

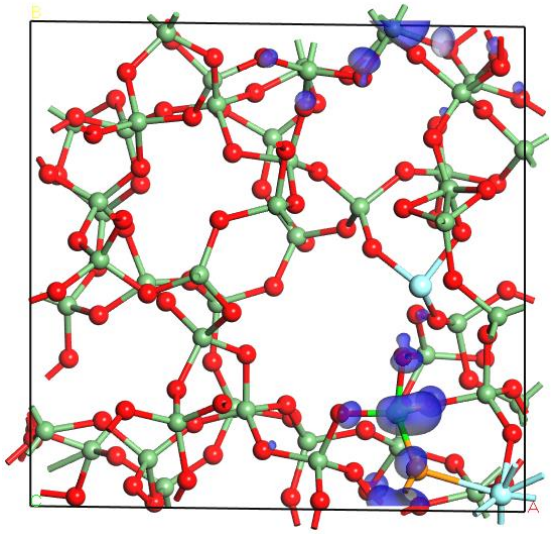


Fig 3.18 (a) Two Y atoms substituted at Ge sites at the VAP in a-GeO_2, (b) PDOS of Y and Ge sites at the VAP in a-GeO_2 calculated by screened exchange; (c) defect orbital at 5.2 eV in the Y substituted VAP.

The process is similar to the passivation of O vacancy gap states by two Y or La atoms in HfO_2 [27,37,38]. This vacancy has a state in the upper gap containing two electrons. Replacing two

Hf's with Y's creates two holes in the valence band. The two electrons previously of the vacancy drop down into the hole states, thereby forming a closed shell.

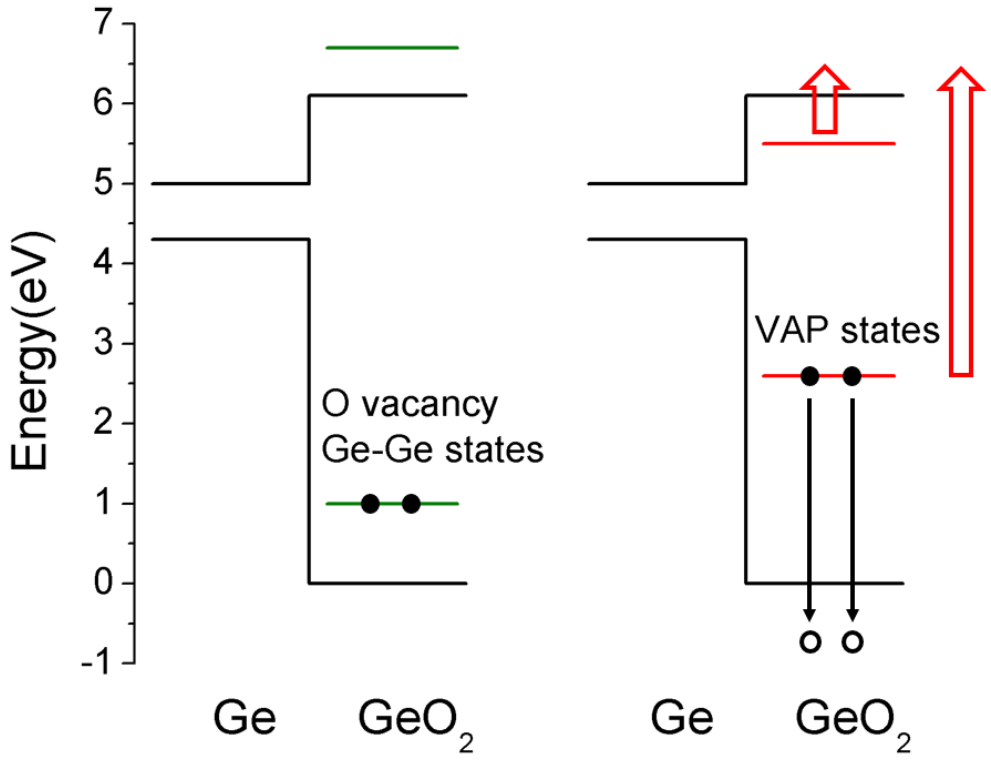


Fig 3.19 Band alignment of Ge and GeO_2, and the defect levels of the O vacancy and VAP. The right figure also summarises the passivation mechanism of the VAP, following the substitution of Ge by Y, with gap states expelled into the GeO_2 conduction band and the fall of two electrons at the Ge1 site into the two valence band holes due to this mechanism.

The role of the VAP is as follows, from **Fig 3.19**. The VAP is not the lowest energy configuration of O deficit in GeO_2, so the O vacancy is the dominant defect of the two. However, the original O vacancy has a filled state only near the GeO_2 VBM. The energy gain from passivating this would only be the 1 eV per electron as they fall from the Ge-Ge bonding state to the VBM. On the other hand, the filled state of a VAP is much higher, and two electrons falling from this state into the hole states at the VBM release much more energy. Thus, the driving force for the passivation is much larger. Thus, the VAP in GeO_2 acts an intermediate state of higher energy than the vacancy itself, that catalyses the transition to the passivated state. However, in SiO_2 the VAP energy is so much higher that it is irrelevant.

We now consider these effects of Y nearer the Ge/GeO_2 interface. There are now a wide range of possible O vacancy and VAP configurations with a range of energies. A reconstructed O vacancy next to the interface is actually an interface with a Ge-Ge bond pointing into the GeO_2

side. Thus, this problem could be re-expressed as the condition for the abruptness of the interface. An O vacancy one further layer into the GeO_2 is the next case. This defect could re-arrange into a VAP separated from the interface, or into one now connected to the interface. Substituting Y for Ge's around these defects gives different possibilities. A critical point is that a direct Y-Ge bond, with the Ge site part of the Ge layer, gives a gap state in the Ge band gap. This is to be avoided. Thus defects where Y substitution of Ge sites on the GeO_2 site is favoured if a beneficial result is desired. We found previously [39] that the advantage of Y (or Sc) over La is that Y favors bonding to O rather than Ge, so that Y segregates into the oxide, whereas La slightly favors bonding to Ge, so that La segregates into the Ge side.

(a) (b)

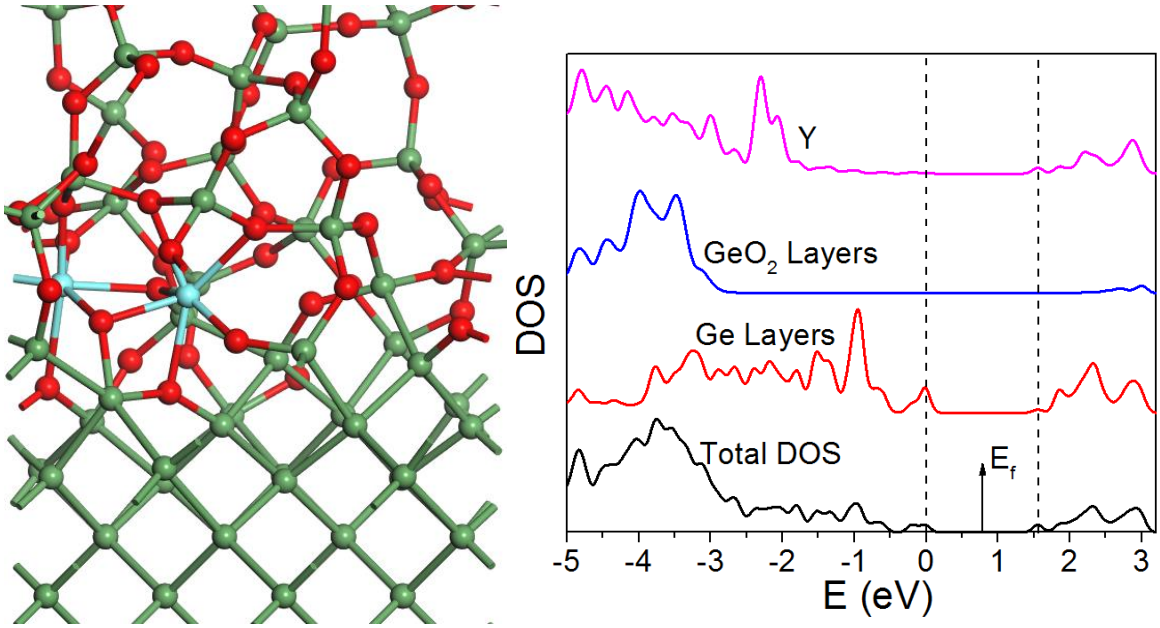


Fig 3.20 (a) Ge/GeO_2 interface where a O vacancy near the interface has been passivated by addition of two Y atoms. Note that they are separated from Ge layer by O bridge bonds. (b) Partial density of states. Note absence of states in the GeO_2 gap.

Many configurations must be checked to give an overall description of the interface case. A useful example is shown in **Fig 3.20**(a,b) which shows two substitutional Y's separated from the Ge side, which were separated by a previous O vacancy. This configuration is fully passivated, with no gap states even in the GeO_2.

4. Conclusions

In conclusion, we have presented DFT calculations on Y doping of GeO_2 to explain the improved reliability of GeO_2 as a gate dielectric on Ge. It is shown that, the VAP type of O

deficiency defect has defect states quite close to the band edges of Ge, which is prone to trap charge carriers. We find that its gap states can be removed by substituting the two Ge atoms near defect site with Y atoms. The successful passivation of defect states is due to electrons localized at gap state falling to the valence holes created by doped Y. This mechanism explains well how Y doping passivates the VAP defects in both crystal and amorphous GeO_2.

5. References

[1] A. Delabie, et al. App. Phys. Lett. **91** 082904 (2007)

[2] S. Takagi, et al. Jpn. J. App. Phys. **54** 06FA01 (2015)

[3] R. Zhang, et al. App. Phys. Lett. **98** 112902 (2011)

[4] R. Zhang, et al. IEEE Trans. Electron Devices **60** 927 (2013)

[5] J. Robertson, et al, Mat. Sci. Eng. R **88** 1 (2015)

[6] G. Groeseneken. et al, Tech. Digest. IEDM (San Francisco, 2014) p34-4.

[7] H. Amimura, et al Tech. Digest. SSDM (Tsukuba, 2016) O-1-01

[8] C. H. Lee, et al. Appl. Phys. Exp. **5** 114001 (2012)

[9] L. Skuja, J. Non-Cryst Solids **149** 77 (1992); M. Fujimaki, et a. Phys. Rev. B **57** 3920 (1998)

[10] S. K. Wang, et al. Jpn. J. App. Phys. **50** 04DA01 (2011)

[11] D. Lencer, et al. Adv. Mater. **23** 2030 (2011)

[12] C. Lu, et al. J. App. Phys. **116** 174103 (2014)

[13] C. Lu, et al. Jpn. J. App. Phys. 021301 (2015); C. Lu, et al. Tech. Digest. IEDM (2015) p14.6

[14] S. Zafar, et al. J. App. Phys. **93** 9298 (2003)

[15] S. J. Clark, et al. Z. Krist **220,** 567 (2005)

[16] S. J. Clark and J. Robertson, Phys. Rev. B **82**, 085208 (2010)

[17] L. Vegard, Zeitschrift für Physik, 5, 17-26 (1921).

[18] J. J. Liang, et al, J Mat Res **14** 11811 (1999)

[19] G. J. Redhammer, et al. Acta Cryst C **63** 193 (2007)

[20] M. Copel, et al. App Phys Lett **97** 182901 (2010)

[21] C. Lu, and A. Toriumi, Tech Digest IEDM (IEEE, 2015) p14.6

[22] C. Lu, et al. App Phys Lett **104** 092909 (2014)

[23] E. Simoen, et al. Mat Sci Semicon Proc **15** 588 (2012)

[24] M. Houssa, et al. App Phys Lett **92** 242101 (2008)

[25] A. Dimoulas, et al. App Phys Lett **96** 012902 (2010)

[26] Y. Tu and J. Tersoff, Phys Rev Lett **89**, 086102 (2002).
[27] L. Lin, K. Xiong, J. Robertson, App. Phys. Lett. **97** 242902 (2010)
[28] E. P. O'Reilly, J. Robertson, Phys. Rev. B **27** 3780 (1983)
[29] M. Kastner, D. Adler, H. Fritzsche, Phys. Rev. Lett. **37** 1504 (1976)
[30] G. Lucovsky, Philos. Mag. B **39** 513 (1979)
[31] K. C. Synder, W. B. Fowler, Phys. Rev. B **48** 13238 (1993)
[32] G. N. Greaves, Philos. Mag. B 37 447 (1978); J. Non-Cryst Solids **32** 295 (1979)
[33] J. Binder, P. Broqvist, A. Pasquarello, App. Phys. Lett. **97** 092903 (2010)
[34] J. F. Binder, P. Broqvist, A. Pasquarello, Physica B **407** 2939 (2012)
[35] H. Li, J. Robertson, Microelec. Eng. **109** 244 (2013)
[36] L. Zhang, et al. ACS App. Mater. Interfaces **8** 19110 (2015)
[37] D. Liu, J. Robertson, App Phys **94** 042904 (2009)
[38] K. Xiong, J. Robertson, S. J. Clark, J. App. Phys. **99** 044105 (2006)
[39] H. Li, J. Robertson, Y. Okuno, J. App. Phys. **120** 134101 (2016)

Chapter 4. GeO_2 Oxidation Mechanism

1. Overview

Silicon has been the dominant semiconductor for many years largely because SiO_2 is such a well-behaved oxide. However, to continue Moore's law scaling, it is becoming necessary to replace Si with a higher mobility semiconductor. Ge has higher electron and hole mobilities than Si and would be a reasonable choice. However, Ge has a poor native oxide GeO_2 with a poorer interface with its parent Ge for reasons that are not fully understood [1-3].

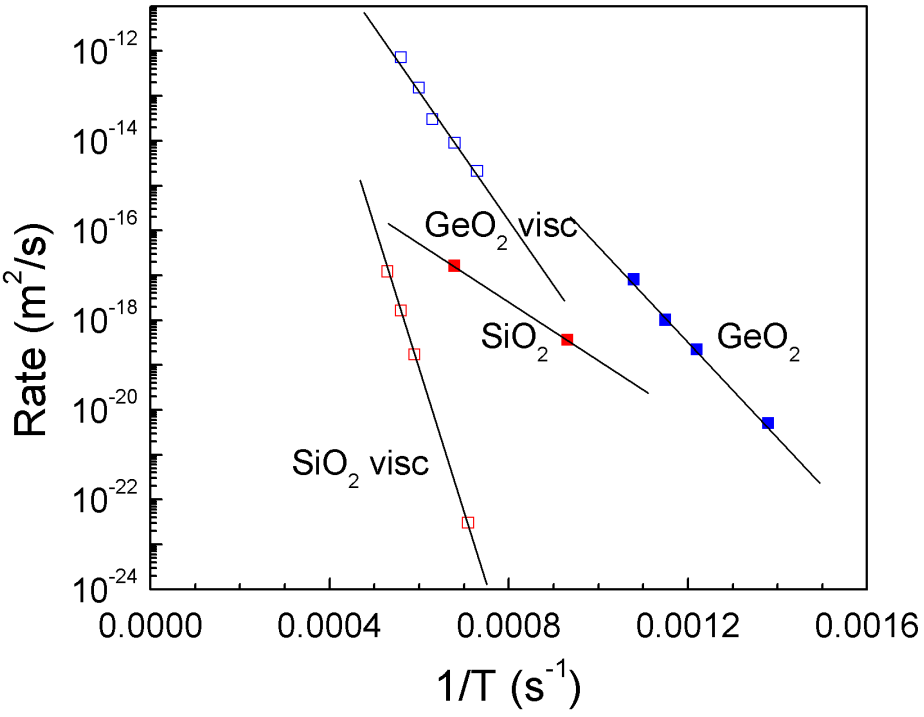
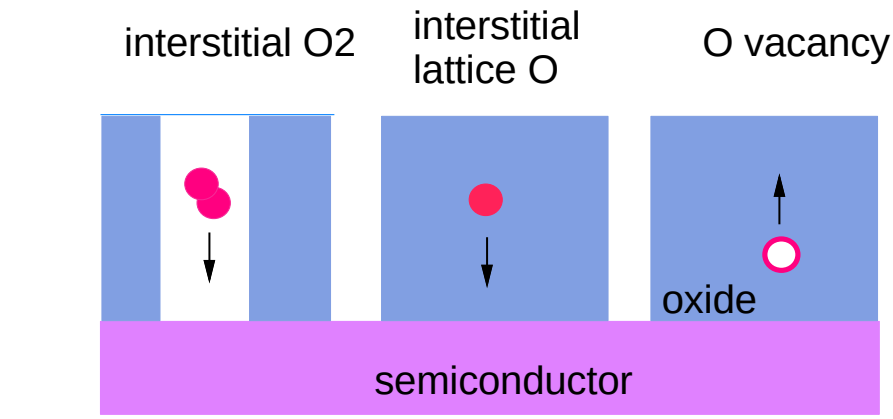


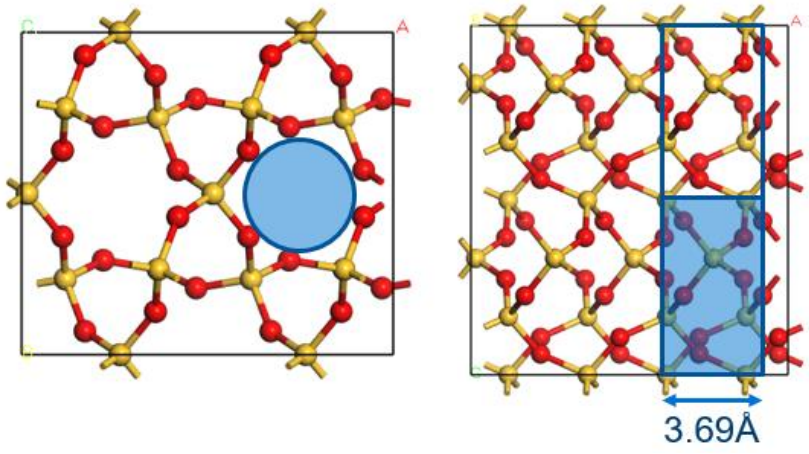
Fig 4.1 Oxidation rates of Ge and Si vs temperature [5,12,13], compared to kinematic viscosity of amorphous GeO_2 and SiO_2 [15-18]. We note that the activation energies for oxidation and viscosity are fairly close for the case of Ge (2.00 eV and 2.74 eV), but quite different for Si (1.23 eV and 5.99 eV), indicating that similar processes may be involved in the case of Ge whereas this is not so for Si.

Historically, one of the notable features of silicon was its well-understood oxidation process. Silicon oxidation follows the linear/quadratic Deal-Grove model [4], in which the O_2 molecule diffuses along interstitial channels of the amorphous a-SiO_2 network [4,5], to react exothermically with Si at the Si/SiO_2 interface [6-9]. This occurs because of the remarkably open network of a-SiO_2. This model was verified by the lack of O isotopic exchange with the existing network oxygens [10,11]. Ge oxidation somehow differs, it creates Ge/GeO_2 interfaces

with more interfacial defects [1,2], molecular GeO is volatile [12], the oxidation kinetics follow an unusual pressure dependence [13], but the Ge/GeO_2 interface can be flat [14]. Here, we analyze the Ge oxidation mechanism in terms of atomic transport processes across the GeO_2 layer, and conclude that it occurs mainly by transport of oxygen network interstitials (also known as peroxyl bridges) rather than by molecular oxygen interstitials [15]. A hint of this is already seen by comparing the experimental oxidation rates of Si and Ge [5,12,13], and the kinematic viscosities of SiO_2 and GeO_2 [16-19] in **Fig 4.1**. The oxidation process happens through the migration of O related defects, which can be limited by the kinetic viscosity of the network. The figure suggests that in GeO_2, the kinetic viscosity and oxidation rate follow similar trend, which is different from the situation in SiO_2. This implies the oxidation mechanism in SiO_2 and GeO_2 could be different.



(a)



(b)

Fig 4.2 (a) Schematic of alternative oxygen transport paths through SiO_2 and GeO_2 to the semiconductor interface. The open circle in the right hand diagram signifies a vacancy. (b) Atomic model of c-SiO_2 network to show the natural channel for O_2 molecules to migrate through.

There are three possible transport paths for the oxidation of Si or Ge, see **Fig 4.2**; diffusion of molecular O_2 interstitials along channels in the oxide, diffusion of interstitial network oxygens, or diffusion of oxygen vacancies across the oxide network. The immobility of the cation in SiO_2 and GeO_2 was confirmed by isotope tracer studies [10,12]. The mechanism for the oxidation is as follows. If Si or Ge is pre-oxidized in ^{16}O and then further oxidized in ^{18}O, if molecular O_2 transport dominates, then an ^{18}O peak will be found at the Si/SiO_2 interface, as indeed occurs for Si [8,9]. If oxidation occurs by transport of O lattice interstitials or vacancies, there is ^{18}O exchange with the network, and a broad peak of ^{18}O will be found across the oxide layer, as is seen for GeO_2 by Xu et al [20] and da Silva et al [21].

2. Calculation setup

Here, we use density-functional supercell calculations to find out which species control the oxidation mechanism in a-GeO_2 by calculating the defect formation energies and the migration energies. The overall diffusion energy is then the sum of these two energies. The calculations are carried out using the CASTEP plane-wave density-functional code [22] and ultra-soft pseudopotentials. The plane-wave cut-off energy is 500 eV. Energies are relaxed until residual forces are below 10^{-5} eV/Å. Some 144-atom and 216-atom GeO_2 network models were made by molecular dynamics, equilibrating at 3000K for 10 ps, then quenching in 10 ps, followed by anenergy minimization. Similar a-SiO_2 networks were created with the same procedure. In addition, supercells of quartz-like SiO_2 and GeO_2 are used. The defect formation energy is the energy cost to form the defect, with the O atom going to form an O_2 molecule outside the SiO_2. The migration energy is the energy barrier of the transition state over which the defect atoms must pass to get to its next site. The total diffusion energy is the sum of these two energies. The migration energies for network defects are derived by the nudged elastic band method [23].

3. Results and Discussions

3.1 O_2* and O lattice interstitial formation energy distribution

Of the three possible diffusing species (Fig. 2), let us first consider the competition between the O_2 molecular interstitial O_2* and the O lattice interstitial O_i. The defect formation energies of O_i and the oxygen vacancy V_O are relatively independent of network, as they depend on the forming or breaking of specific covalent bonds. In contrast, the formation energy of the O_2* arises from the closed-shell repulsions of the O_2 molecule and the surrounding oxide network, and thus it depends on the size of the network interstice. The O_2* molecule is surrounded by lattice oxygens. In quartz SiO_2 of density 2.66 gm/cm^3, there is one main interstice

configuration, and the defect formation energy at this site is ~2.04 eV, **Fig 4.3**, similar to the results found by Bongiorno [6] but larger than in Hamann [24].

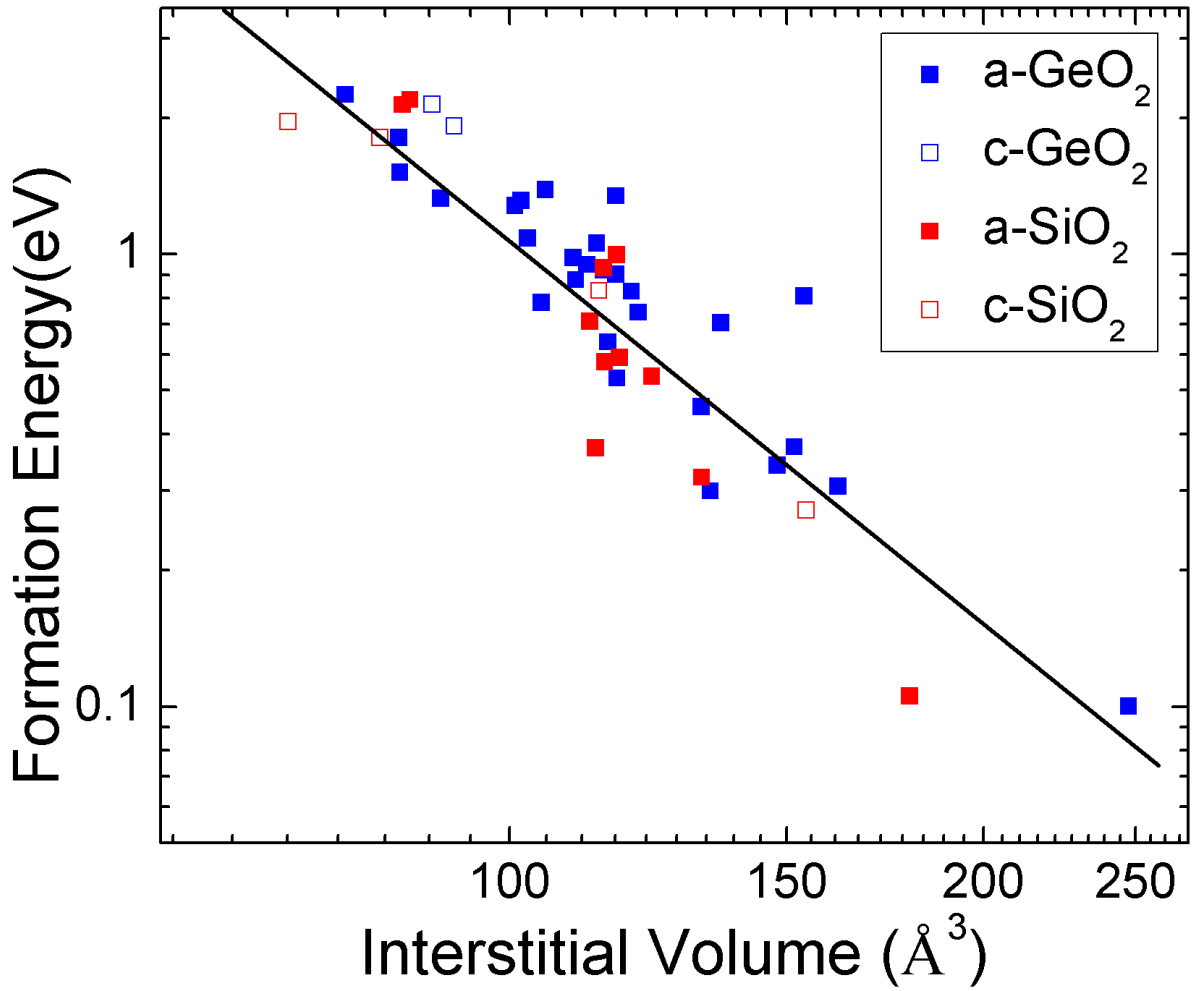


Fig 4.3 Formation energy vs relaxed interstitial volume of O_2* in SiO_2 and GeO_2.

A-SiO_2 (silica) is less dense than quartz at 2.25 gm/cm^3 and it contains a range of interstices and channels mostly of larger size [25,26]. The lower density means that the Si-O-Si bond angle increases from 144^0 in quartz towards 150^0 in silica. **Fig 4.3** plots the O_2* formation energy vs the relaxed interstice volume. This energy varies inversely with volume and follows a $V^{-3.5}$ dependence. The interstice volume is calculated from the distance between the O_2 bond center to the six nearest network O sites, averaged, and then approximated as a sphere. A similar calculation is carried out for quartz-like GeO_2 and for a-GeO_2 random networks whose density is 3.55 gm/cm^3 [27]. **Fig 4.3** shows that the O_2* defect formation energy follows the same line as in SiO_2, confirming that the energy depends mainly on O-O repulsions in both cases.

We studied many interstices in a-SiO_2 and a-GeO_2 in both their initial unrelaxed and relaxed state (i.e. before and after the O_2 molecule is added). **Fig 4.4** shows the relaxation ratio (relaxed volume/unrelaxed volume) vs. interstitial volume for each oxide. A major difference between a-SiO_2 and a-GeO_2 is that a-SiO_2 has a longer tail of large interstitial volumes. This arises from the larger O bond angle and different network topology of silica giving fewer low-order rings. This allows a greater interstice relaxation for the SiO_2 case.

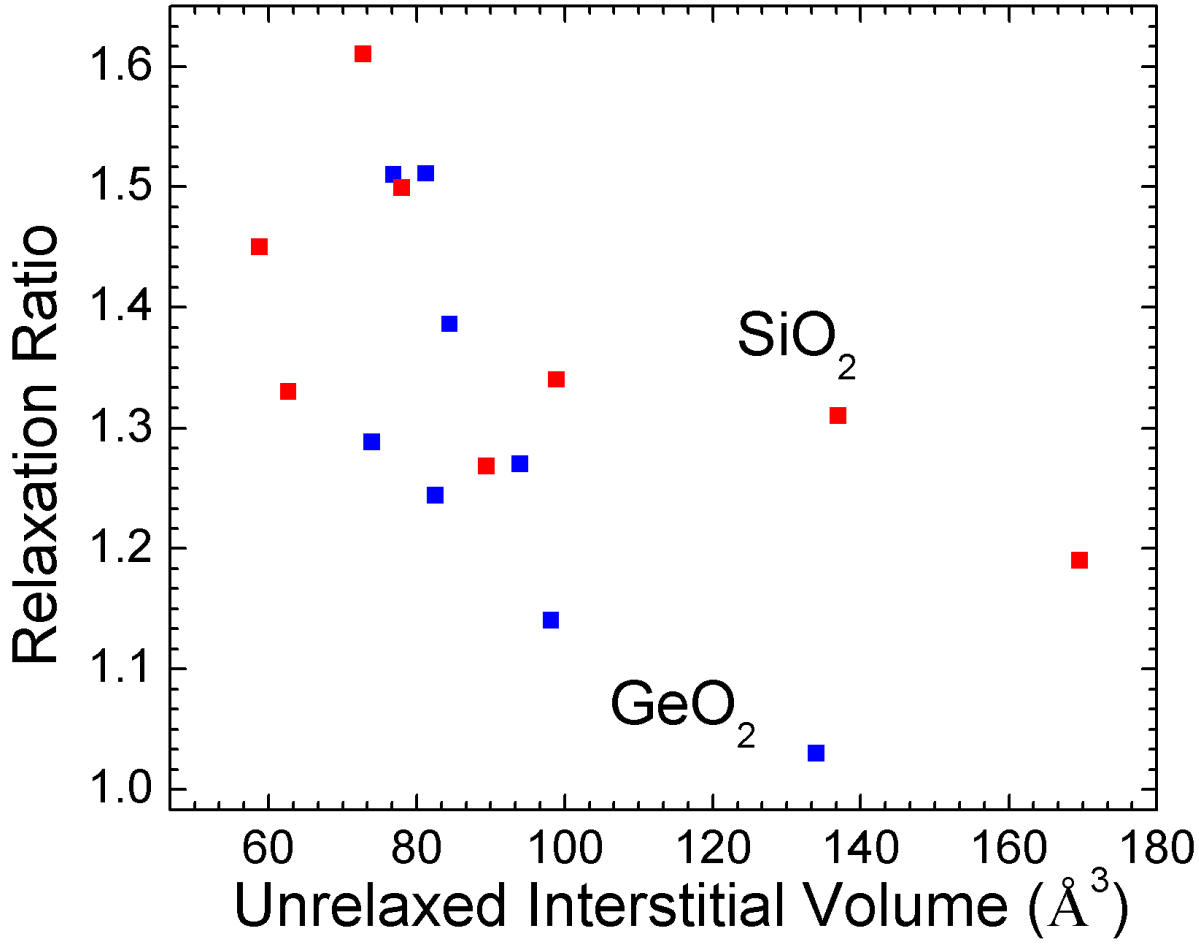


Fig 4.4 Relaxation ratio vs initial unrelaxed interstital volume for a-SiO_2 and a-GeO_2, noting larger relaxation for the a-SiO_2 case. Blue point for GeO_2 and red point for SiO_2.

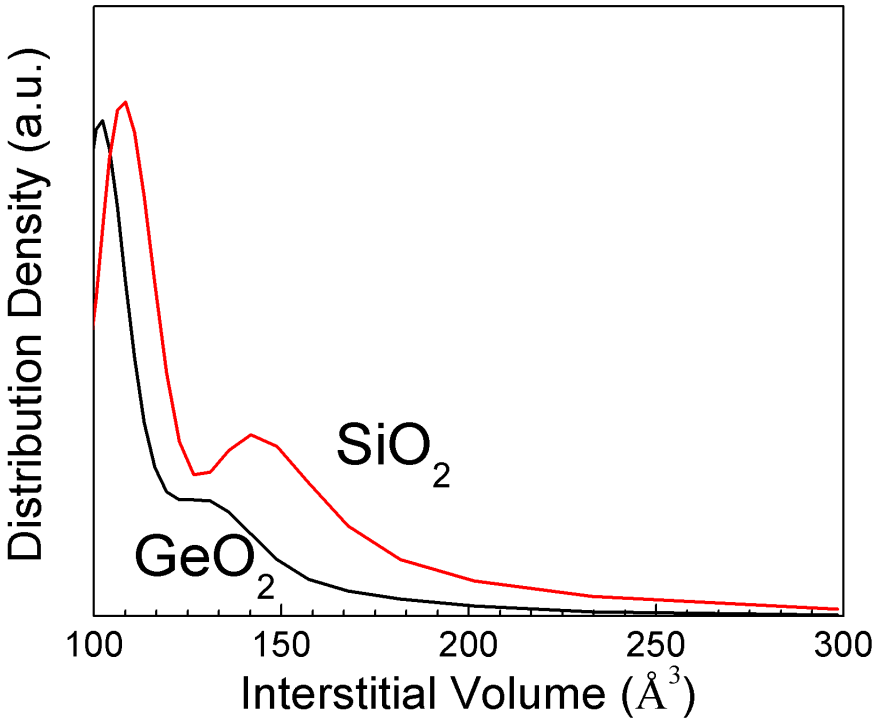


Fig 4.5 Probability distribution of relaxed interstitial volumes of O_2* in a-SiO_2 and a-GeO_2 network models, guassian broadened by 5Å^3

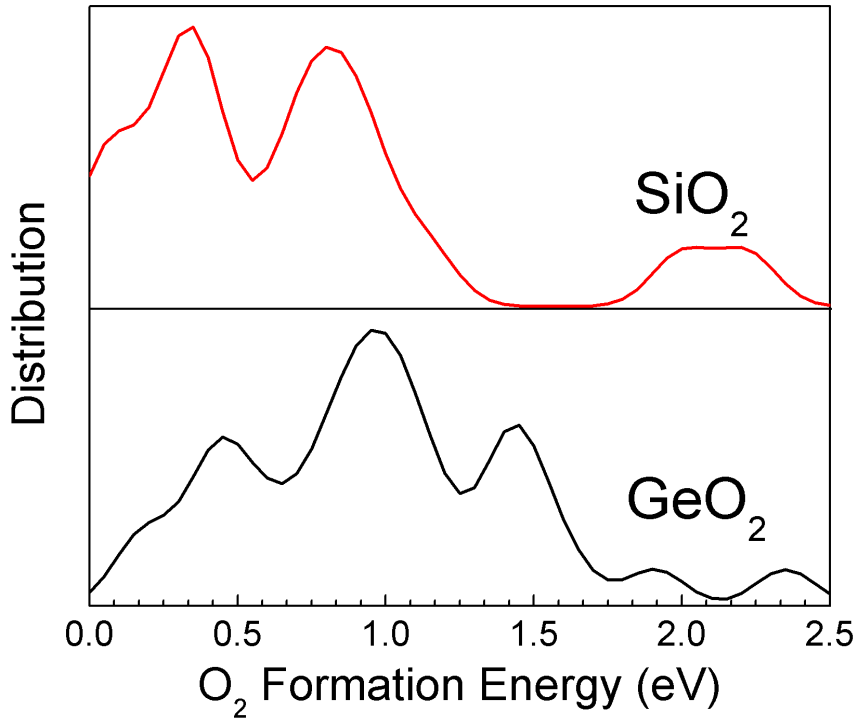


Fig 4.6 Probability distribution of relaxed interstitial formation energy of O_2* in a-SiO_2 and a-GeO_2 network models, guassian broadened by 0.2eV.

Fig 4.5 plots the probability distribution of relaxed interstitial volumes for both a-SiO_2 and a-GeO_2. Interestingly, the average volume is rather similar, around 110 $Å^3$, and slightly larger for the SiO_2 case. Thus, from the interstitial volumes in **Fig 4.5** and the relationship in **Fig 4.3**, the O_2* defect formation energy in a-GeO_2 is expected to be lower in a-SiO_2.

Fig 4.6 plots the probability distribution of relaxed interstitial formation energy for both a-SiO_2 and a-GeO_2. Interestingly, in a-GeO_2 the O_2 formation energy distribution has got a fatter tail in the high energy range compared to that in a-SiO_2. Thus, from **Fig 4.6**, the O_2* defect formation energy is indeed lower in a-SiO_2, which is consistent with our analysis on the interstitial volume. In addition, the formation energy of O_2* is notably smaller in a-SiO_2 than in quartz.

3.2 O_2* and O lattice interstitial diffusion

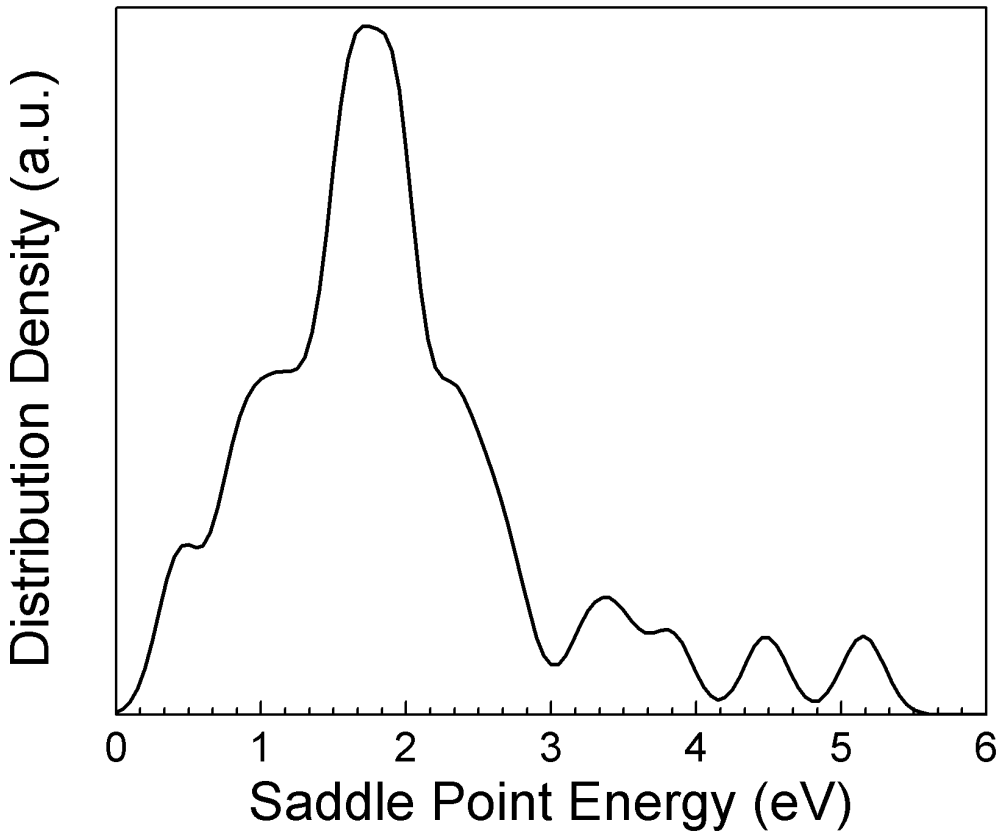


Fig 4.7 Probability distribution of diffusion barrier energies for a-GeO_2 models.

The O_2 or O diffusion energy is the sum of the defect formation energy and the migration (barrier) energy. For the O_2* in an interstice site in quartz, there are two diffusion paths, along the O_x channel or along the $O_{x,y}$ channels with quite different diffusion barriers. In silica or a-GeO_2, there are a range of diffusion paths. As in Biongorno and Pasquarello [6,7], there are roughly six adjacent interstitial sites around each O_2* site for it to hop to. Thus, diffusion can

be simplified to a percolation problem on a simple cubic mesh, for which the percolation threshold is 0.55 as previously used [6]. Thus, the overall diffusion energy is the energy at 55% of its cumulative distribution. To have a good estimation of the migration barrier and diffusion energy, we have calculated 30 independent diffusion process for O_2*, applied Gaussian widen smearing to get the distribution and take the 55% cumulative distribution as the threshold, as shown in **Fig 4.7** The formation energy, migration barrier, and diffusion energy of various O defects in a-GeO_2 and a-SiO_2 are summarized in **Table 4.1**.

	a-GeO_2			a-SiO_2		
Bond angle	130^0			$144\text{-}150^0$		
Bulk heat of formation/O	**3.02**			**4.85**		
Heat of formation (exp) [31]	**3.06**			**4.78**		
	Formation energy (eV)	Migration barrier (eV)	Diffusion energy (eV)	Formation energy (eV)	Migration barrier (eV)	Diffusion energy (eV)
O_2 interstitial / O_2	0.97	0.95	1.92	0.7	0.7	**1.40**
O interstitial$^{\$}$	0.42	1.37	**1.79**	1.40	1.39	2.79
O vacancy $^{\$}$	3.16	2.54	5.7	6.05	4.28	10.33
O vacancy$^{\#}$	0.14	2.54	2.68	1.20	4.28	5.48
VAP$^{\#}$	0.82	1.71	2.53	2.2	3.5	5.7
Experiment [12,5]			**2.00**			**1.23**

Table 4.1 Calculated defect formation energies and migration barrier energies in GeO_2 and SiO_2. $ in O rich-limit, # in O-poor limit. The experimental diffusion energies in the last row are to be compared with the specific calculated values in bold above.

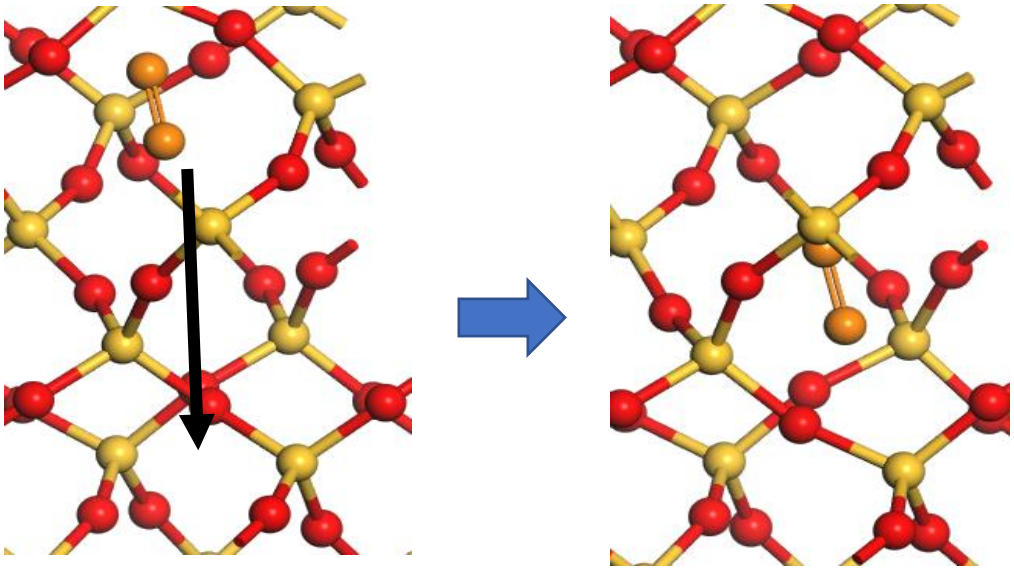


Fig 4.8 Migration paths of O_2* in quartz SiO_2

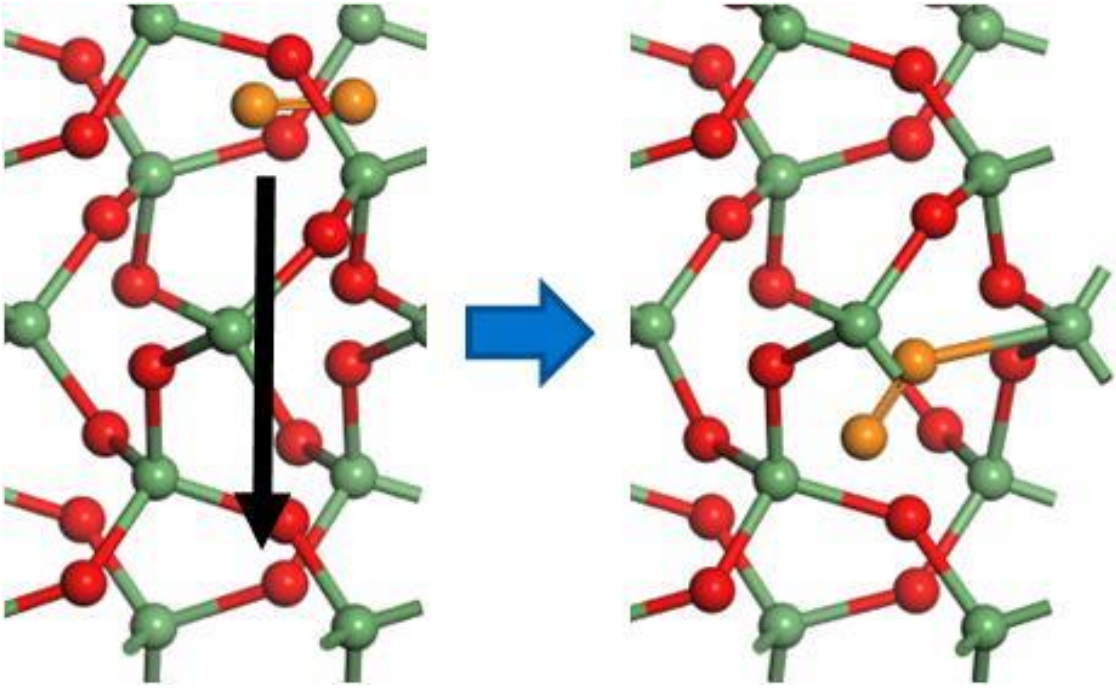


Fig 4.9 Migration paths of O_2* in quartz-like GeO_2 showing slight bonding to channel in the transition state

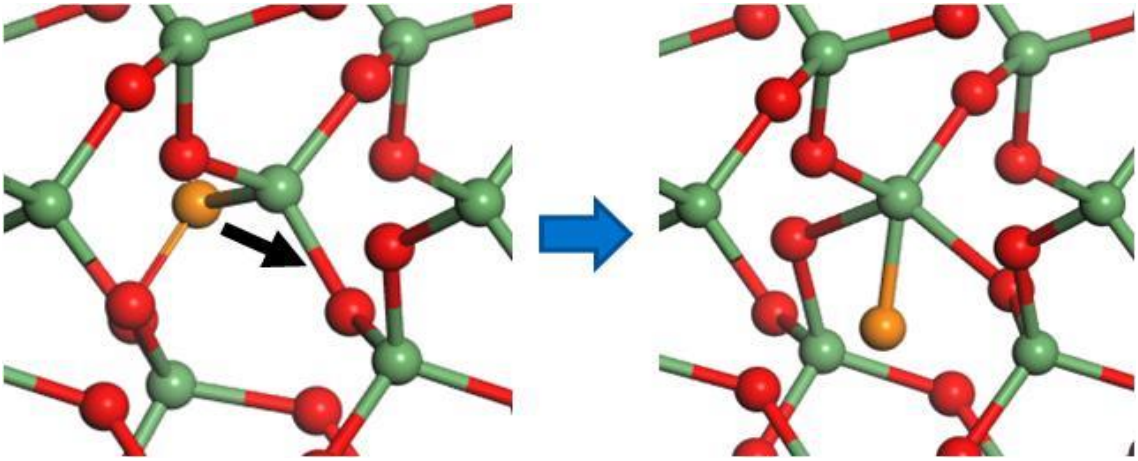


Fig 4.10 Migration path of oxygen network interstitial in SiO_2 and GeO_2 via 5-fold intermediate Si/Ge site.

Our calculation shows that the O_2* migration energy being larger in a-GeO_2 than in a-SiO_2, see **Table 4.1**. This is because the migration channels between interstices can dilate more in a-SiO_2 during the passage of O_2* giving it a lower migration barrier. To better illustrate this point, we have also calculated the diffusion of O_2* in c-SiO_2 and a-GeO_2, as shown in **Fig 4.8**, and **Fig 4.9**. c-GeO_2 provides narrower diffusion channel for O_2* than c-SiO_2 because of the bonding angle difference. During diffusion process, the O_2* molecule also forms temporary bonds with lattice Ge due to smaller lattice deformation endurance. Similar effects can be found in

amorphous structures which makes O_2* more difficult to diffuse in a-GeO_2 than in a-SiO_2. This gives a total diffusion energy of ~1.4 eV for a-SiO_2 and ~2.0 eV for a-GeO_2 for O_2* as in **Table 4.1**. The key difference for a-GeO_2 is its smaller O bond angle, which raises its O_2* migration barrier, due to its reduced channel relaxation. (We also note the greater interstitial volume in silica than quartz because of its lower mass density, which increases the O_2* diffusion rate by five orders magnitude at 600^0C [26].)

We now consider the neutral O lattice interstitial O_i or 'peroxyl bridge'. Its configuration is shown in **Fig 4.10**. The oxygen bond angles are typically 100-110^0. We see in **Table 4.1** that the formation energy of O_i changes from 2.0 eV in quartz to 1.4 eV in a-SiO_2 to ~0.4 eV in a-GeO_2, for O-rich chemical potentials. In a-GeO_2, the defect formation energy has a distribution. It depends somewhat on the Ge-Ge separation of the initial Ge-O-Ge unit, as this determines the two bond angles at the oxygens. The O_i migrates from one Si-O-Si or Ge-O-Ge bridge to an adjacent bridge site via a 5-fold coordinated Si or Ge transition state in both oxides, as shown in **Fig 4.10**. A similar transition state was seen by Hamann [24].

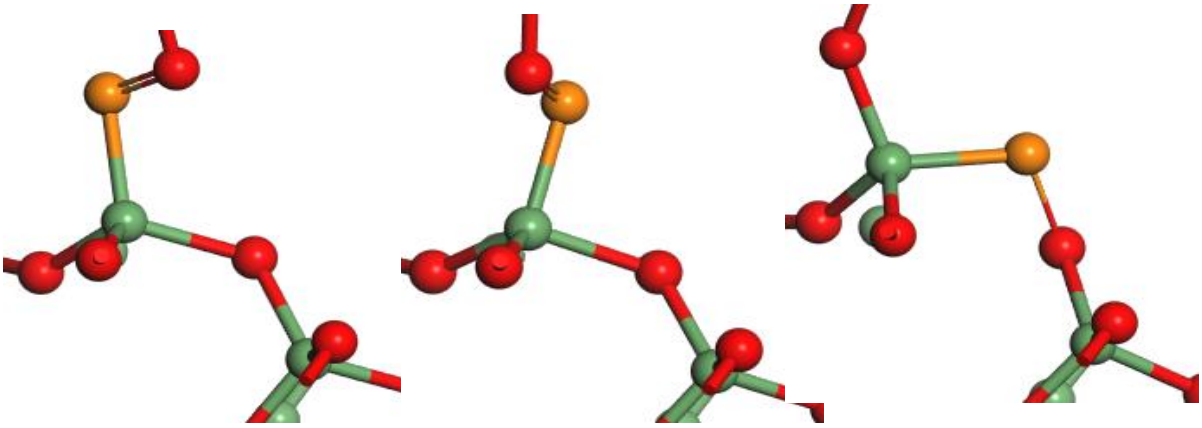


Fig 4.11 Unusual altenative low energy, low probability path of O_i in a-GeO_2.

We have also found some unusual alternative low diffusion barrier path for O_i in a-GeO_2. In these paths, the transition state geometry is the distorted O_i defect (all of the bonds are kept) rather than a 5-fold Ge. Such low diffusion barrier is attribtued to no bonds breaking during the diffusion process.

The overall effect is that the lowest energy diffusion path for oxygen in silica is the O_2* route because of its open channels, whereas the lowest path in a-GeO_2 is by the network interstitial O_i. Thus, O diffuses by the O_2* route in a-SiO_2, giving an activation of oxidation of the O_2* diffusion energy 1.4 eV, while in a-GeO_2 the O_2* diffusion energy is increased, and the process shifts to the O_i path which has a diffusion energy of ~1.8 eV. These values are compared to

experiment in **Table 4.1** (in bold). The O_2* diffusion energy of 1.4 eV in a-SiO_2 is consistent with the experimental activation energy of 1.23 eV for Si oxidation in **Fig 4.1**, and the O_i diffusion energy of 1.79 eV for a-GeO_2 is consistent with the experimental activation energy of 2.0 eV for Ge oxidation.

3.3 O vacancy and VAP defect formation and diffusion energy

The oxygen vacancy is one of the most important defect in GeO_2. It exists in the form of Ge-Ge bonds. The formation energy of oxygen vacancy is quite low at oxygen poor condition, thus the oxygen vacancy concentration is usually high at Ge:GeO_2 interface. Especially, in this interface region, there usually exist several oxygen deficient GeO_x layers. The typical oxygen vacancy diffusion is shown in **Fig 4.12**. The Ge dimer bond at the oxygen vacancy site breaks and forms two 3-fold Ge atoms. An O atom bonded to one of the 3-fold Ge atom moves across, has the bonding angle distorted, and finally fills the original oxygen vacancy site by forming two new Ge-O bonds. A new oxygen vacancy is formed at the original place of this moving O atom. We have calculated plenty of oxygen vacancy diffusion process to determine the formation energy and migration barrier. It turns out to be that oxygen vacancy has both formation energy and migration barrier in GeO_2 much lower than it is in SiO_2, as shown in **Table 4.1**. This could be explained by the fact that Ge-O bonds are weaker than Si-O bonds, thus it is easier to break the Ge-O bonds.

A final detail about defects is the valence alternation pair (VAP). This consists of a positively charged 3-fold O site, and a negatively charged 3-fold Ge site. These sites have been seen in simulations of a-GeO_2 and a-SiO_2 [28,29]. Briefly, they are O-deficiency defects created from O vacancies. The Ge-Ge dimer bond of the oxygen vacancy bond breaks, and one of the Ge atom move across to form a dative bond with a nearby O atoms, leaving the other Ge atom 3-fold. Their formation energy is greater than the vacancy, but their migration energy is much lower than for the vacancy, Table 4.1. Thus, the total diffusion barrier is similar to the oxygen vacancy, which makes them relevant to the oxidation processes in these oxides.

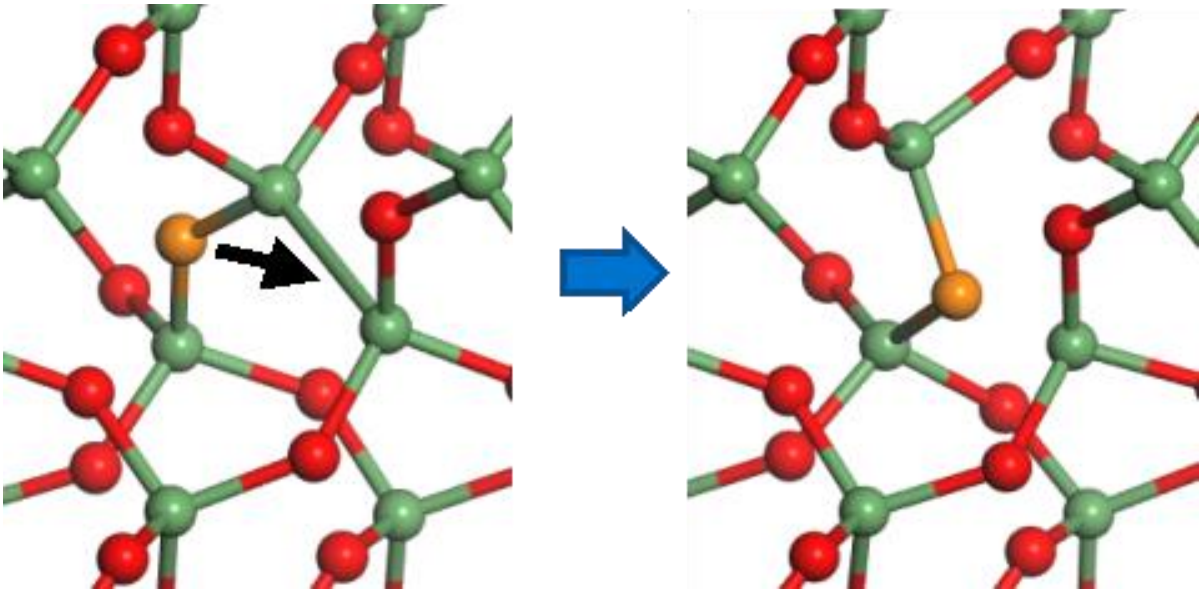


Fig 4.12 Migration path of O vacancy (equivalent to rebond Ge-Ge bond) to adjacent site.

3.4 Competition between O deficiency defect and O excess defect.

We finally consider competition between V_O and O interstitials. The O vacancy is a very costly defect in SiO_2, so this is not involved in Si oxidation (except in any reactive layer next to the interface [30]). On the other hand, the O vacancy is thought to be important in a-GeO2, and has been considered to be a key diffusing species [9,16]. This is because the O vacancy has a low cost in GeO_2 in the O-poor condition, as occurs at the Ge/GeO_2 interface. However, the O-rich condition dominates across most of the film under oxidation conditions.

To understand the relevance of the interstitials and vacancies for diffusion in detail, we note that defect formation energies depend on the relative oxygen chemical potential, μ(O), which is difference between oxygen chemical potential and the chemical potential for O_2. μ(O) varies from μ(O) = 0 eV at atmospheric pressure at the film's surface to μ(O) = -4.85 eV at the Si/SiO_2 interface, where –4.85 eV is the bulk heat of formation of SiO_2 per O atom. Or, in GeO_2, μ(O) varies from 0 eV at the external surface to μ(O) = -3.02 eV at the Ge/GeO_2 interface, where -3.02 eV is the bulk heat of formation of GeO_2 per O. The defect formation energy E_{form} at an arbitrary O chemical potential μ1 is given by

$$E_{form}(\mu_1) = E_{form}(\mu_0)+n*\mu_1 \quad (1)$$

Where $E_{form}(\mu_0)$ is E_{form} at μ(O) = 0 eV, and n=+1 for vacancies and n=-1 for interstitials.

At the SiO_2 surface where μ(O)=0 eV, O_2 has a fairly low formation energy and V_O has a very large formation energy of 6.05 eV. At the Si/SiO_2 interface, V_O is now a lower cost defect, but

it still costs 1.4 eV. Therefore, O_2 diffusion dominate most SiO_2 region and is the key component of silicon oxidation. On the other hand, GeO_2 is different. At the GeO_2 surface, O_i is low cost, at ~0.5 eV and V_O costs 3.16 eV. However, at the Ge/GeO_2 interface V_O is very low cost at only 0.14 eV, much lower than at Si/SiO_2. Thus, O_i and V_o jointly transport oxygen from upper surface to the Ge:GeO_2 interface to help germanium oxidation. Under diffusing conditions, the oxygen flux J across a layer is given by

$$J = -\sum_i D_i \frac{\partial c_i}{\partial x} = -\sum_i D_i \frac{\partial c_i}{\partial \mu} \cdot \frac{\partial \mu}{\partial x} \qquad (2)$$

where the concentration of species i: $c_i = c_0 \exp(-E_{form,i}/kT)$. The flux must be the same across the whole film. This equation can describe two situations for GeO_2. In condition 1, the quiescent state, the low cost of V_O at the Ge/GeO_2 interface dominates, and the μ(O) of O-poor condition spreads across most of the film to near the surface, and then changes quickly to the O-rich condition very close to the surface, Fig.13. μ(O) changes from fully O-rich to O-poor across the film thickness, and the main change occurs over a fraction ($kT/\Delta\mu$) of this thickness below the surface.

In condition 2, corresponding to oxidation, the O-rich condition spreads from the GeO_2 surface down through most of the film to the Ge/GeO_2 interface. This causes the O flux to be carried by the O interstitial. μ(O) then rapidly falls very close to the interface, **Fig 4.13**, V_O then becomes the flux carrier, formally across the last atomic layer, and the sharp decrease in μ(O) occurs to create a large concentration gradient to force the same flux to be carried by V_O.

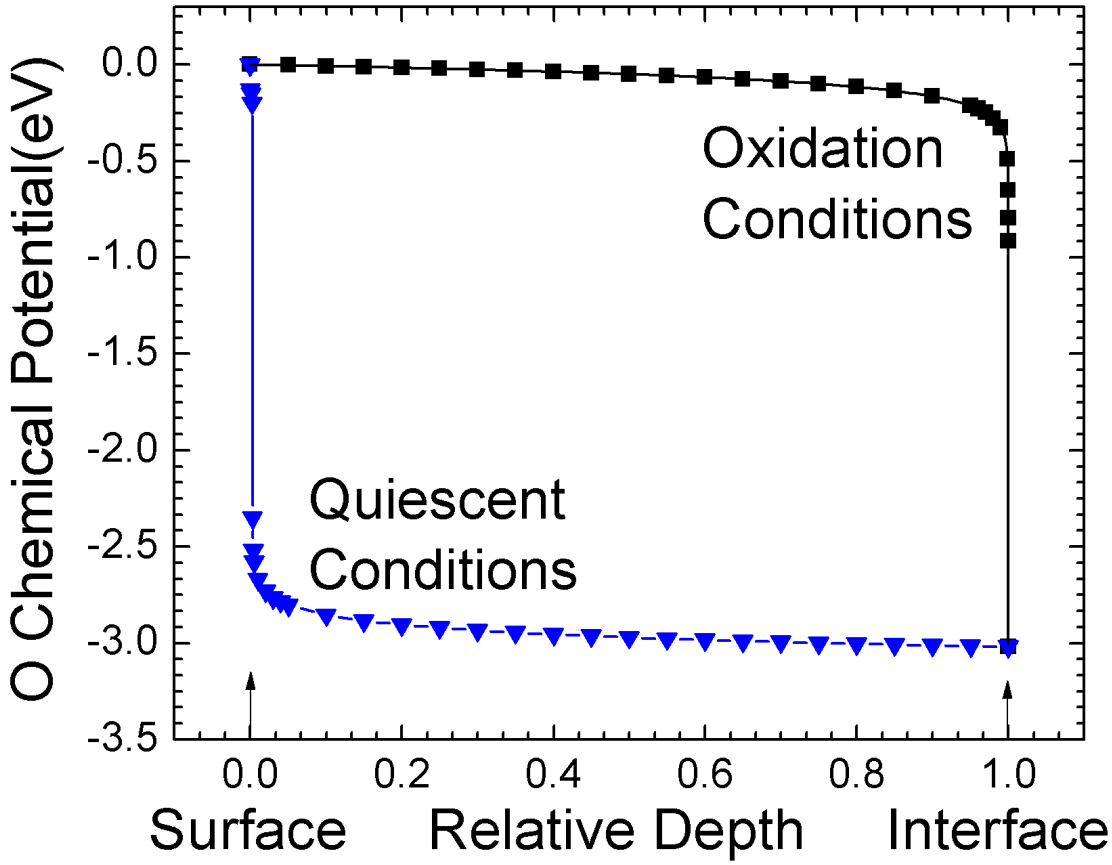


Fig 4.13. O chemical potential vs depth across a GeO_2 film while in contact with Ge substrate and while oxidising.

The difference of V_O between SiO_2 and GeO_2 arises as follows. Making a V_O leaves two dangling bonds on the adjacent Si or Ge sites. These sites rebond into Si-Si or Ge-Ge bonds. The formation energy of an O vacancy is very large in SiO_2 even after the rebonding (6.05 eV) in the O rich condition, because of the large cost of breaking Si-O bonds. It is less in GeO_2 (3.16 eV) because the Ge-O bond is weaker. (Table 4.1). Furthermore, the migration barrier for the O vacancy is very high in SiO_2, at 4.28 eV, and is much less in GeO_2, at 1.71 eV. The migration energy largely consists of undoing the reconstruction of the vacancy into a Si-Si or Ge-Ge bond.

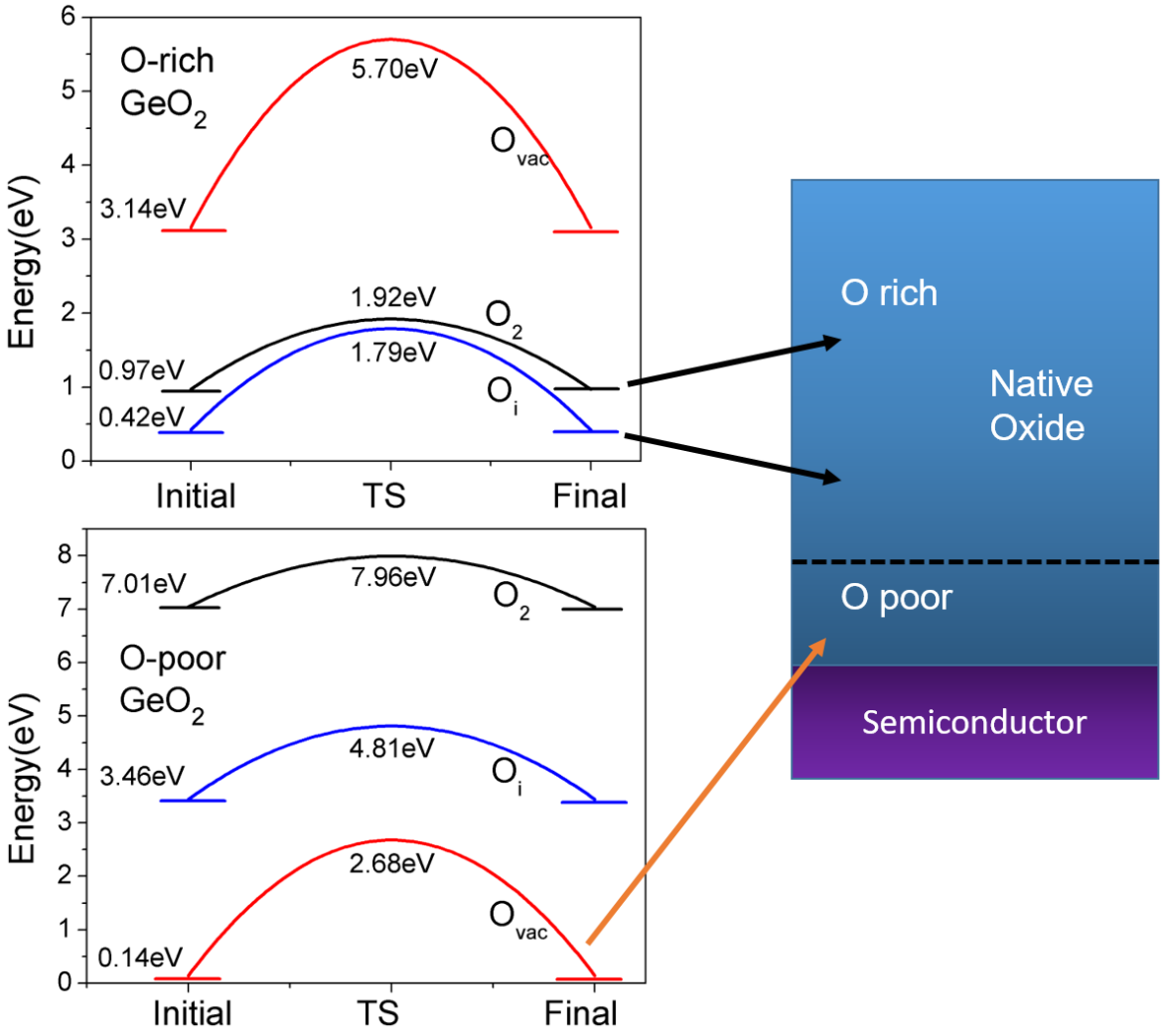


Fig 4.14. Different diffusion barrier energies of O_2*, O_I and V_O at high μ(O) (surface) and at low μ(O) (interface), for SiO_2 and GeO_2 compared, and their relevance to transport mechanisms at different region of the GeO_2 film (right).

The overall germanium oxidation process diagram is shown in **Fig 4.4**. In the upper part of GeO_2 layers, the chemical environment is close to the O rich condition since it is in contact with air. At O rich condition, the O excess defects (O_i and O_2*) has lower formation energy, which makes O_i the defect with the lowest diffusion barrier. Therefore, O_i diffusion dominates the O transporation in the upper region of GeO_2 layers. On the other hand, in the lower part of GeO_2 layers, the chemical environment is close to the O poor condition since it is in contact with Ge layers. At O poor condition, the O deficiency defects (oxygen vacancy and VAP) has lower formation energy, which makes oxygen vacancy the defect with the lowest diffusion barrier. Thus, oxygen vacancy diffusion dominates the O tranporation process instead of O_i.

Such analysis is consistent with the isotope experimental results, in which Si or Ge is pre-oxidized in ^{16}O and then further oxidized in ^{18}O. Our simulation shows that molecular O_2 transport dominates in SiO_2, thus there should be nearly no O exchange with the SiO_2 network, which explains why an ^{18}O peak was found at the Si/SiO_2 interface[8,9]. On the other hand, germanium oxidation occurs by transport of O lattice interstitials or vacancies, and thus there is ^{18}O exchange with the network. This explains the finding of a broad peak of ^{18}O across the oxide layers, as is seen for GeO_2 by Xu et al [20] and da Silva et al [21].

4. Conclusions

In summary, the main diffusing species changes from the interstitial O_2* molecule in SiO_2 to the lattice O interstitial in a-GeO_2. This is largely due to the smaller O bond angle in GeO_2, meaning fewer high order rings, and this less ability for its network to dilate around the O_2* to allow it to lower its migration energy, unlike in a-SiO_2. This is consistent with isotopic tracer results [10,12]. The O lattice interstitial dominates O O vacancy diffusion in a-GeO_2 because this becomes O-poor only very close to the Ge/GeO_2 interface under oxidizing conditions.

5. References

[1] Y. Kamata, Materials Today **11** 30 (2008)

[2] A. Delabie, F. Bellenger, M. Houssa, T. Conard, S. VanElshocht, M. Caymax, M. Heyns, M. Meuris, Appl. Phys. Lett. **91** 082904 (2007)

[3] J. Robertson, R. M. Wallace, Mat. Sci. Eng. R **88** 1 (2015)

[4] B. E. Deal, A. S. Grove, J. Appl. Phys. **36** 3770 (1965)

[5] J. D. Plummer, M. D. Deal, P. B. Griffin, 'Silicon VLSI Microelectronics', Prentice Hall (2000) p319

[6] A. Bongiorno, A. Pasquarello, Phys. Rev. Lett. **88** 125901 (2002)

[7] A. Bongiorno, A. Pasquarello, Phys. Rev. B **70** 195312 (2004)

[8] A. Bongiorno, A. Pasquarello, Phys. Rev. Lett. **93** 086102 (2004); J. Phys. Conden. Mat. **17** S2051 (2005)

[9] L. Tsetseris, S. T. Pantelides, Phys. Rev. Lett. **97** 116101 (2006)

[10] F. Rochet, S. Rigo, M. Froment, C. D'Anterroches, C. Maillot, H. Roulet, G. Dufour, Adv. Phys. **35** 237 (1986)

[11] E. P. Gusev, H. C. Lu, T. Gustafsson, E. Garfunkel, Phys. Rev. B **52** 1759 (1995)

[12] S. K. Wang, K. Kita, T. Nishimura, K. Nagashio, A. Toriumi, Jpn. J. App. Phys. **50** 04DA01 (2011)

[13] C. H. Lee, T. Tabata, T. Nishimura, K. Nagashio, A. Toriumi, App. Phys. Exp. **5** 114001 (2012)

[14] T. Sasada, Y. Nakakita, M. Takenaka, S. Takagi, J. App. Phys. **106** 073716 (2009)

[15] G. Hetherington, K. H. Jack, J. C. Kennedy, Phys. Chem. Glasses **5** 130 (1964)

[16] G. Urbain, Y. Bottinga, P. Richelt, Geochim. Cosmochim. Acta **46** 1061 (1982)

[17] E. F. Riebling, J. Chem. Phys. **39** 1889 (1963)

[18] S. K. Sharma, D. Virgo, I. Kushiro, J. Non-Cryst. Solids **33** 235 (1979)

[19] W. Xu, T Nishimura, T. Yajima, A. Toriumi, presented at JSAP meeting (March 2017, Yokohama, 16p-413-4)

[20] S. R. M. daSilva, G. K. Rolm, G. V. Soares, I. J. R. Baumvol, C. Krug, L. Miotti, F. L. Freire, M. E. H. M. daCosta, C. Radike, App. Phys. Lett. **100** 191907 (2012)

[21] H. Li, J. Robertson, IEEE-SISC (San Diego, Dec 2016) p7.2

[22] S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. J. Probert, K. Refson, M. C. Payne, Z Krist **220,** 567 (2005); S. J. Clark and J. Robertson, Phys. Rev. B **82**, 085208 (2010)

[23] G. Henkelman, H. Jónsson, J. Chem. Phys. **113**, 9901 (2000)

[24] D. R. Hamann, Phys. Rev. Lett. 81 3447 (1998)

[25] S. L. Chan, S. R. Elliott, Phys. Rev. B **43** 4423 (1991)

[26] R. A. B. Devine, J. Capponi, J. Arndt, Phys. Rev. B **35** 770 (1987)

[27] M. Micoulaut, L. Cormier, G. S. Henderson, J. Phys. Conden. Mat. **18** R753 (2006)

[28] H. Li, J. Robertson, App. Phys. Lett. **110** 032903 (2017);

[29] A. Pasquarello, M. S. Hybertsen, R. Car, Nature 396 58 (1994)

[30] R. M. C. deAlmeida, S. Goncalves, I. J. R. Baumvol, F. C. Stedile, Phys. Rev. B **61** 12992 (2000)

[31] O. Kubaschewski, C. B. Alcock, 'Metallurgical Thermochemistry', (Pergamon, Oxford, UK, 1979)

Chapter 5. Reduced Fermi Level Pinning on Ge by Silicides/Germanides

1. Overview

For both conventional semiconductors and layered semiconductors, the performance of their devices is often limited by their contact resistances [1]. This is due to the Schottky barriers at the contacts. A standard way to minimize the Schottky barrier heights (SBH) is to vary the contact metal, as the SBH will vary with the metal work function. However, this is difficult for many semiconductors as they suffer from Fermi level pinning, in which the SBH varies only weakly with the work function. This is expressed in terms of the Fermi level pinning factor $S = \partial\varphi_n/\partial\Phi_M$ being very small, where φ_n is the n-type barrier height and Φ_M is the work function of the contact metal. For Si, the Fermi level pinning problem and a large SBH can be circumvented by heavily doping the Si, so that carriers can tunnel through the depletion layer, but this is less easy for other semiconductors where dopants are less soluble or not always shallow.

For example, germanium has higher electron and hole mobilities than Si and it is one of the more promising candidates for a next generation channel material. However, for Ge the n-type SBH is very large because the Fermi level is pinned close to its valence band edge [2,3]. Many approaches have been tried to overcome this problem, such as introducing ultra-thin oxide layers [4], or delta-doping with As, S or Cl [5]. Another method is to use germanides or silicides [6-12] which appear to have weaker Fermi level pinning [13].

Schottky barrier behaviors are often classified as being limited by intrinsic properties [14-17] or by extrinsic effects such as interfacial defects [18]. Generally, in Si or Ge the intrinsic effects tend to dominate. The intrinsic states include the metal induced gap states (MIGS) [14]. The MIGS are extensions of the travelling wave states of the metal continued into the band gap of the semiconductor and are drawn from the valence and conduction states of the bulk semiconductor [14,17]. As such, in a cubic semiconductor, the pinning factor S should be a function only of the semiconductor itself and *independent* of the type of metal, while the pinning energy or charge neutrality level (CNL) should be independent of the crystal face. Generally, if extrinsic effects are present, such as the interfacial defects at the contacts on MoS_2, these tend to reduce S below the intrinsic MIGS value [19-21] and so they are viewed as a negative factor.

However, there is a second type of 'extrinsic factor' in which varying the type of metal and the face can also be used to vary the Schottky barrier height. If this extrinsic quality can be accessed in a suitable way, it might be used to reduce the SBH of n-Ge.

It was previously noted that the SBHs of silicides on Si were much more weakly pinned than those of elemental metals [13] – that is, S is much larger. Recently Nishimura et al [22] found that the SBH of germanides on Ge were also more weakly pinned than for elemental metals. They also found that their SBHs on Ge were face-dependent, whereas the SBHs of elemental metals were independent of face [3]. This is reminiscent of the behavior of $NiSi_2$ where some years ago Tung [23] found that the $Si/NiSi_2(111)$ interface had very different SBHs for its A and B orientations. Thus, here we carry out a detailed study of the face dependence of Schottky barrier heights of different metals.

2. Methods

We have calculated the SBHs of metal silicides and germanides on Si or Ge for metals of a wide range of work functions, using supercell models of the semiconductor and the silicide or germanide. The calculations are carried out using the plane wave density functional code CASTEP [24,25]. We use the generalized gradient approximation for the electron exchange correlation functional, and norm-conserving pseudopotentials with a cutoff energy of 750 eV. The density of states calculations for Ge used a different pseudopotential which gives a band gap even in GGA, generated by OPIUM code. The convergence is carried out to an energy below 10^{-5} eV per atom, and with forces below 10^{-3} eV/Å. A k-point mesh of 4x4x2 is used for Brillouin zone integrations. The calculations are carried out on supercells of 5 layers of silicide and 9 layers of Si (and no vacuum layer). The lattice geometry is relaxed in GGA. When there is a lattice mismatch between the Si and the silicide, the x,y lattice constants of the silicide are kept fixed to those of the Si and the vertical z distances are allowed to relax. This roughly conserves the silicide volume. The work function of a metal depends mostly on its atomic volume.

The GGA functional is known to under-estimate the semiconductor band gap, thus it will also under-estimate the Schottky barrier height, as has been known for some years [26]. Here we can correct this effect by also calculating the density of states using the screened exchange (sX) hybrid density functional [25]. Generally though, we are interested in *changes* in SBH due to the face or rates of change of SBH with work function, which are not affected by the absolute band gap.

3. Results

3.1 Si/$NiSi_2$ interface structure

For silicide metals of higher work function, we use the cubic $NiSi_2$ structure as a representative structure, as this has a high symmetry and is lattice-matched to Si. The Si/$NiSi_2$ interface is interesting because the covalent bonding of the Si continues across the interface into the silicide. In $NiSi_2$, the Ni sites are 8-fold coordinated by Si, and Si sites are 4-fold coordinated by Ni [23,27]. The bonds in $NiSi_2$ themselves are only weakly polar [13]. At the interface, the $NiSi_2$ lattice terminates in 7-fold Ni sites, which can be viewed as Ni 'dangling bonds', while the Si sites all remain 4-fold bonded across the interface [27], **Fig 5.1(a)**. The Ni can be replaced by other transition metals to form their silicides. This $NiSi_2$ structure is used for Ti, Cr, Fe, Ru, Co, Ni and Pt.

The (100) interface of $NiSi_2$ has been much less studied than (111). Its structure was first modeled as containing 6-fold coordinated Ni sites [28], but it was then noticed that it possessed a 2x1 reconstruction [29]. On the basis of total energy calculations, Yu et al [30] proposed an unusual 2x1 structure (**Fig 5.1(b)**) which was later confirmed by Falke et al [31] using high resolution scanning transmission electron microscopy (STEM). The 2x1 reconstructed interface has 5-fold Si sites and lateral Si-Si bonds on the Si side.

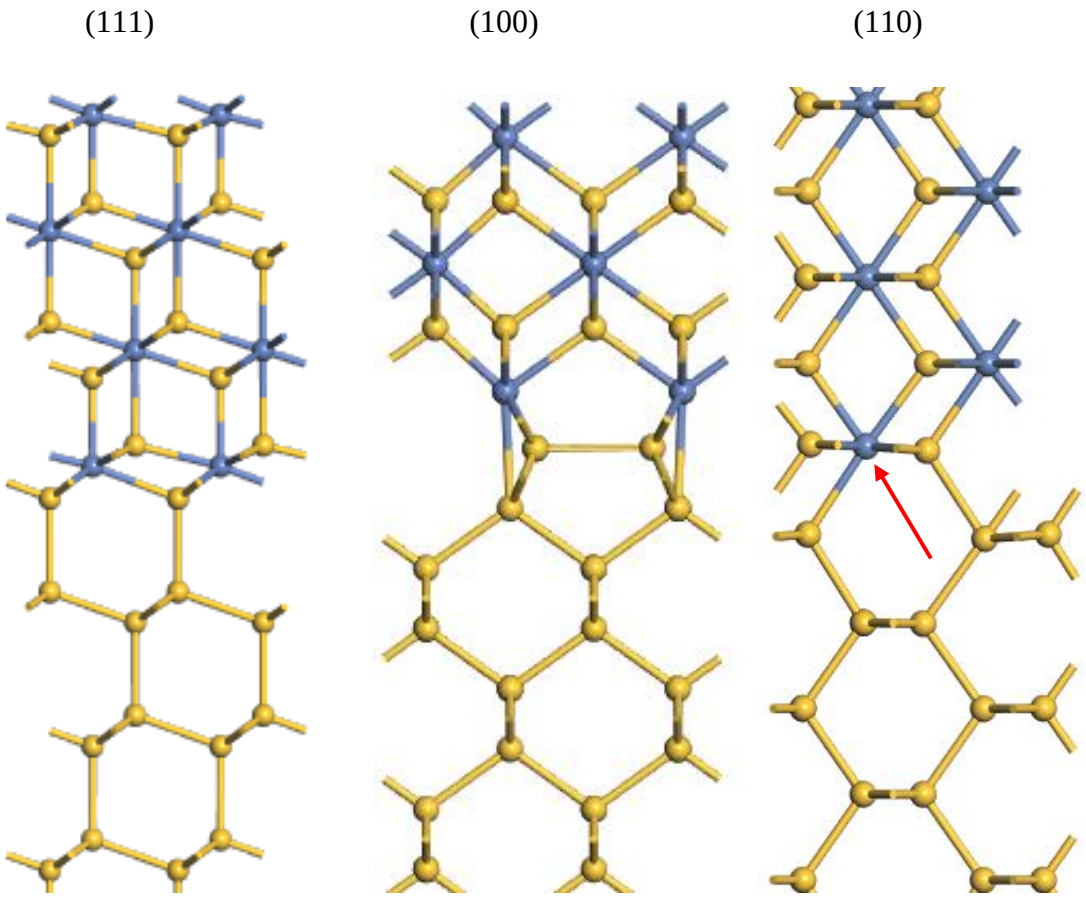


Fig 5.1 The atomic structure of Si/$NiSi_2$ interfaces for different Si surface facet (a) (111), (b) (100), and (c) (110). The red arrow shows the dangling bond on metal atom.

3.2 Si/YSi_2 interface structure

For silicides of metals of larger atomic radius, we use the YSi_2 structure. This has a hexagonal unit cell which is lattice-matched to the $\sqrt{3}x\sqrt{3}(111)$ face of Si [11], **Fig 5.2(a)**. Here, the silicons lie in a plane and the Y forms layers above and below the silicons. The bonding in this lattice is very polar. The Si-terminated (0001) face of YSi_2 bonds to Si(111). While YSi_2 is hexagonal, its *a* and *c* lattice constants are quite similar. This allows us to construct a pseudo-(100)Si/YSi_2 interface by rotating the YSi_2 lattice to give the structure shown in **Fig 5.2(b)**. This structure is used for the silicides of Yb, Y, La and Zr. Thus, using $NiSi_2$ and YSi_2 lattices, we are able to study silicide/germanide interfaces from Yb to Pt, covering the widest range of metal work functions.

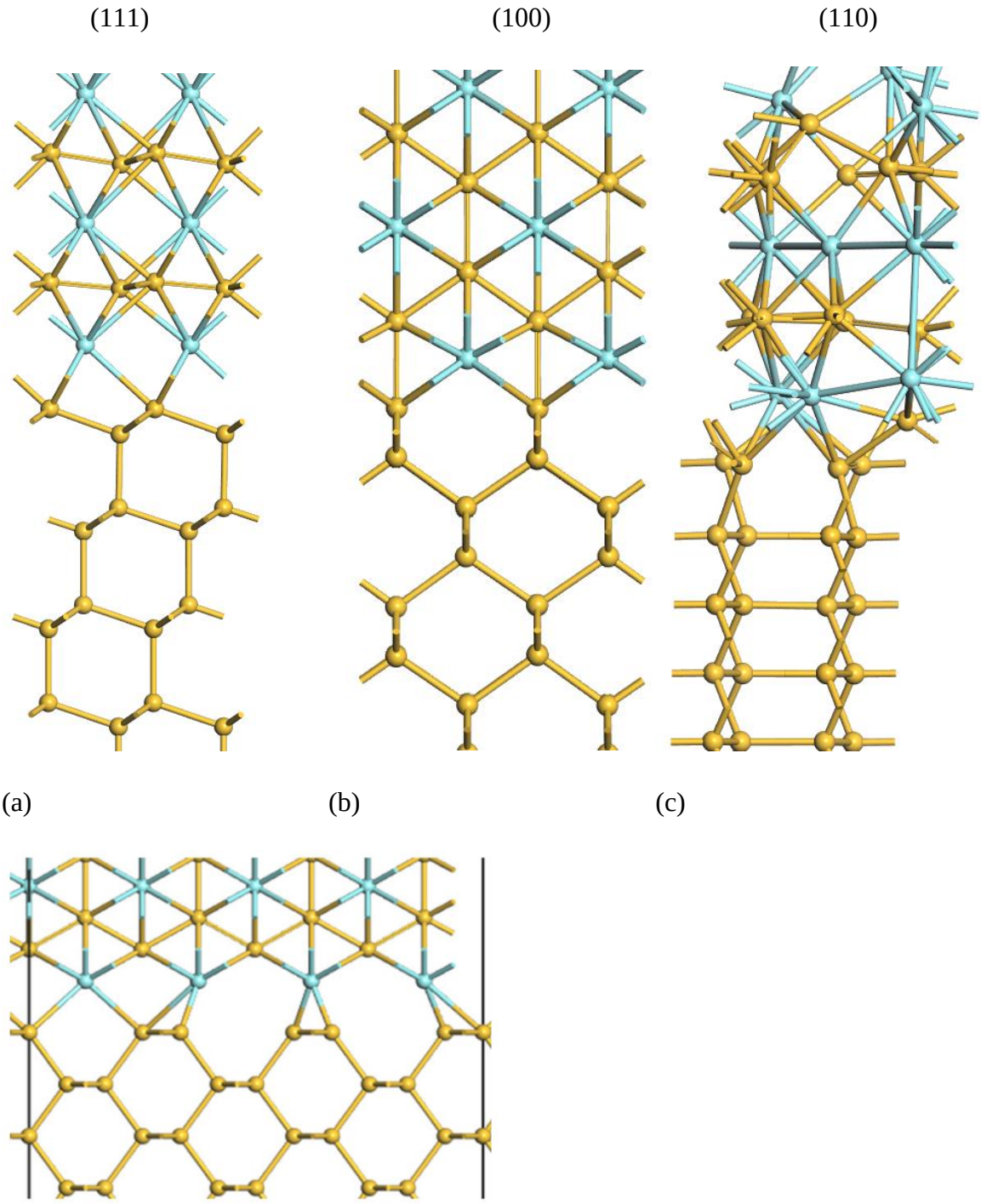


Fig 5.2 Atomic structure of (a) (111), (b) (100), and (c) (110) Si/YSi_2 interface, the (110) interface has got disordered YSi_2 layers for better lattice matching. (d) The large (110) Si/YSi_2 interface, not used in calculation due to model size.

The structure of the (110)Si:$NiSi_2$ interface is unknown, but a simple coherent interface model shown in **Fig 5.1(c)** is used to represent this interface. We also built a (110) interface of YSi_2 using the lattice matched interface shown in **Fig 5.2(d)**. However, this model has too large a lateral size and leaves too many dangling bonds on the Si side, and thus has a large interfacial energy. Instead, we took a slab of (110)YSi_2 and disordered it by a molecular dynamics quench. This was then attached to the (110) face of Si and structurally relaxed. The resulting interface is shown in **Fig 5.2(c)**. It has no Si dangling bonds. This allowed us to study (110)Si/MSi_2 interfaces for the full range of silicides using a combination of the $NiSi_2$ and YSi_2 structures. The resulting SBHs for the two types showed a continuous line of SBH values and reasonably well-behaved barrier heights.

3.3 Schottky barrier height for Si/silicides interface

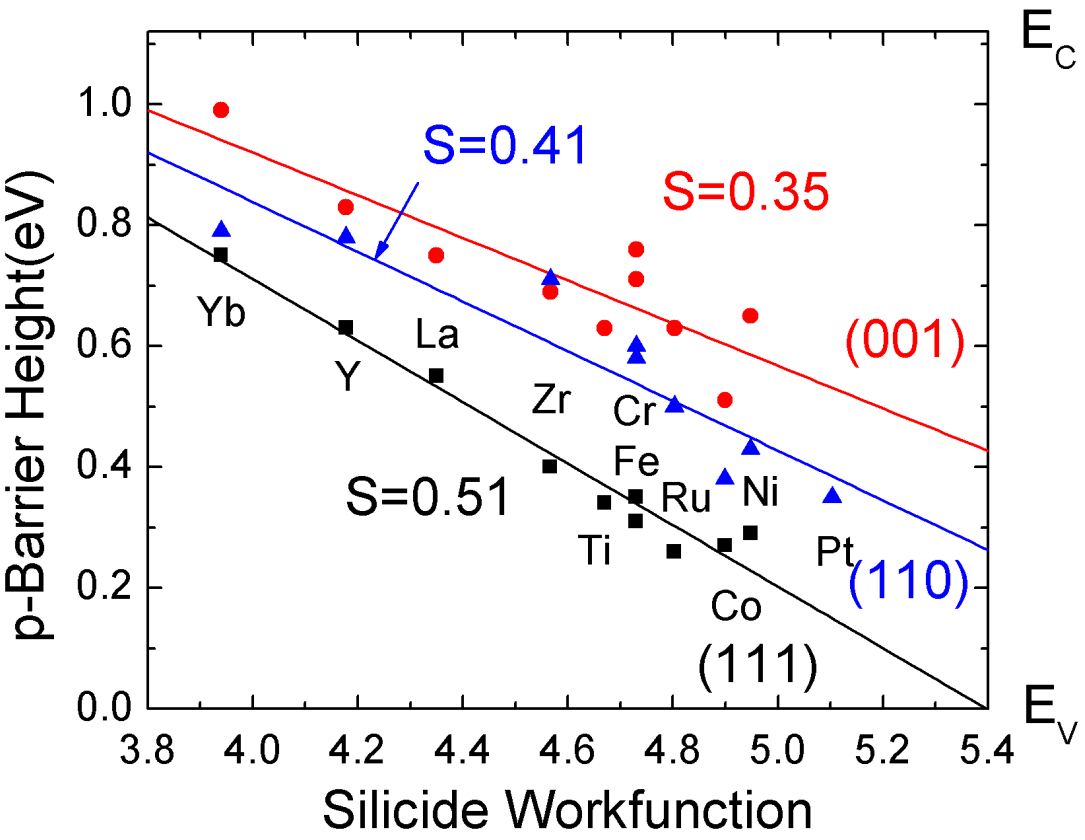


Fig 5.3 Calculated p-type Schottky barrier heights for Si/metal silicides, for (111), (110) and (100) interfaces, by the density functional supercell method.

The φ_n values for each face are plotted against the silicide work function in **Fig 5.3**. This work function is estimated from the work function of the parent metal [32] using a Miedema average [6], WF = $(\Phi_M \Phi_{Si}^2)^{1/3}$. We see that the calculated φ_n values in **Fig 5.3** lie on separate lines for the (100) and (111) faces. We note the continuity of data points for $NiSi_2$ and YSi_2 lattices

indicates that the φ_n values depend on work function and not the precise crystal structure. The (100) values are offset upwards from the (111) points. The calculated slope is S = 0.35 for (100) faces, and S = 0.51 for (111) faces. The slope for the most complicated interface (110) is found to be 0.41.

On average, the slopes are large, of order S~0.4, compared to those of elemental metals which are very small (S~ 0.05). This shows that silicide SBHs have much weaker Fermi level pinning than the SBHs of elemental metals. The calculated slopes for silicides are consistent with the experimental data [6,10].

3.4 Schottky barrier heights for Ge/germanides interface

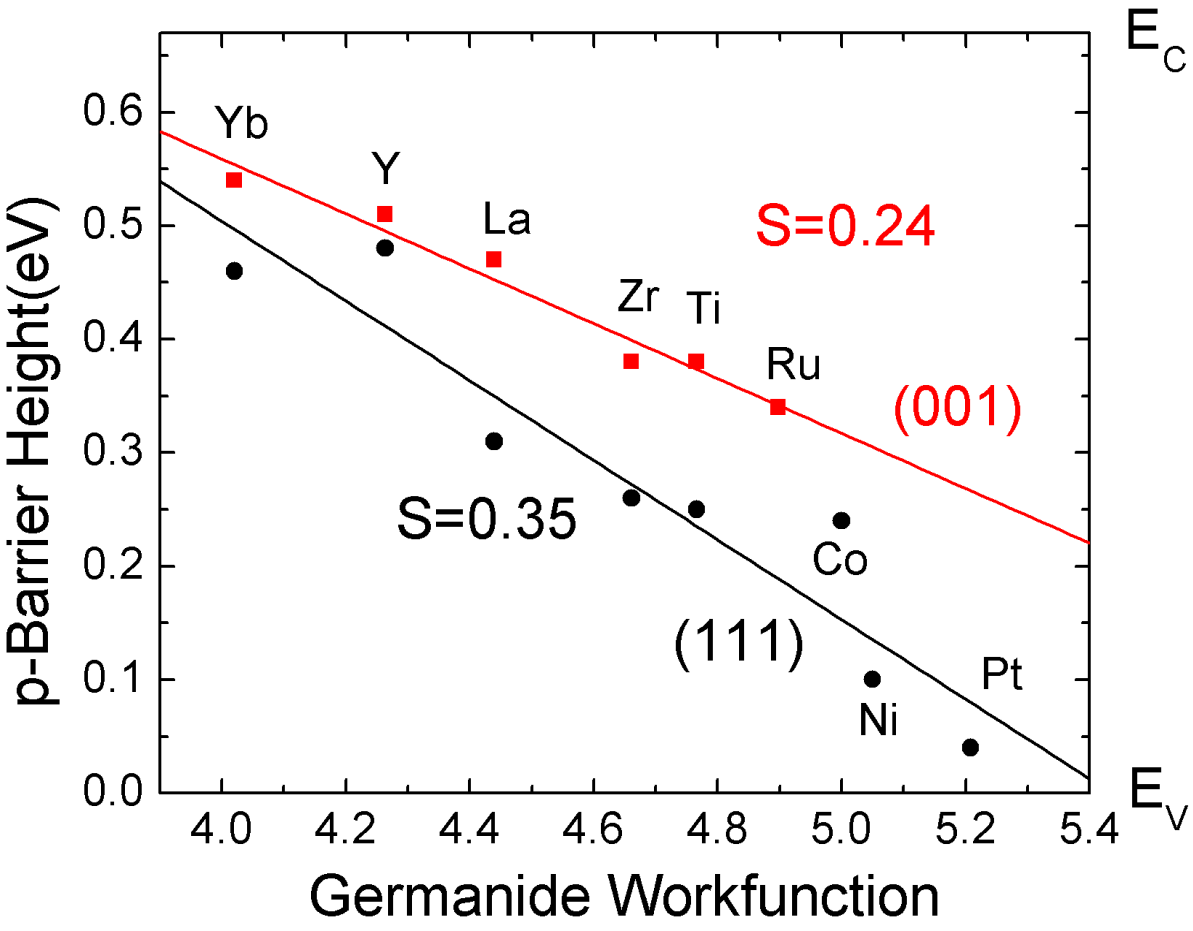


Fig 5.4 Calculated p-type Schottky barrier heights for Ge/metal germanides, for (111), and (100) interfaces, by the density functional supercell method.

A similar set of calculations were carried out for germanides on Ge. Due to the narrower band gap of Ge, and the band gap under-estimation of local density formalism, it is more difficult to locate the band edges of Ge, and there is greater uncertainty in the SBH values for Ge systems. Nevertheless, we see that there is a similar behavior for the Ge systems. A slope

of 0.24 is found for the (100) face of germanides and a slope of 0.35 for the (111) faces, **Fig 5.4**. This means that the barrier heights of germanides also show weaker Fermi level pinning, like the silicides. This gives rise to smaller SBH values for the electropositive germanides on n-type Ge, as in experiment [17]. It is this large slope factor which allows the dependence of SBH on face to occur for silicides and germanides. They are both an 'extrinsic effect'. Our calculations [13] accounted for the wellknown change in SBH from the Si/$NiSi_2$(111)A to (111)B interface orientation seen by Tung [23]. They now explain the less well-known but much larger (0.4 eV) reduction in n-type SBH from Si/$NiSi_2$ (111)A to (100) also seen by Tung [33,34].

3.5 Comparison of DFT Schottky barrier heights with experimental values.

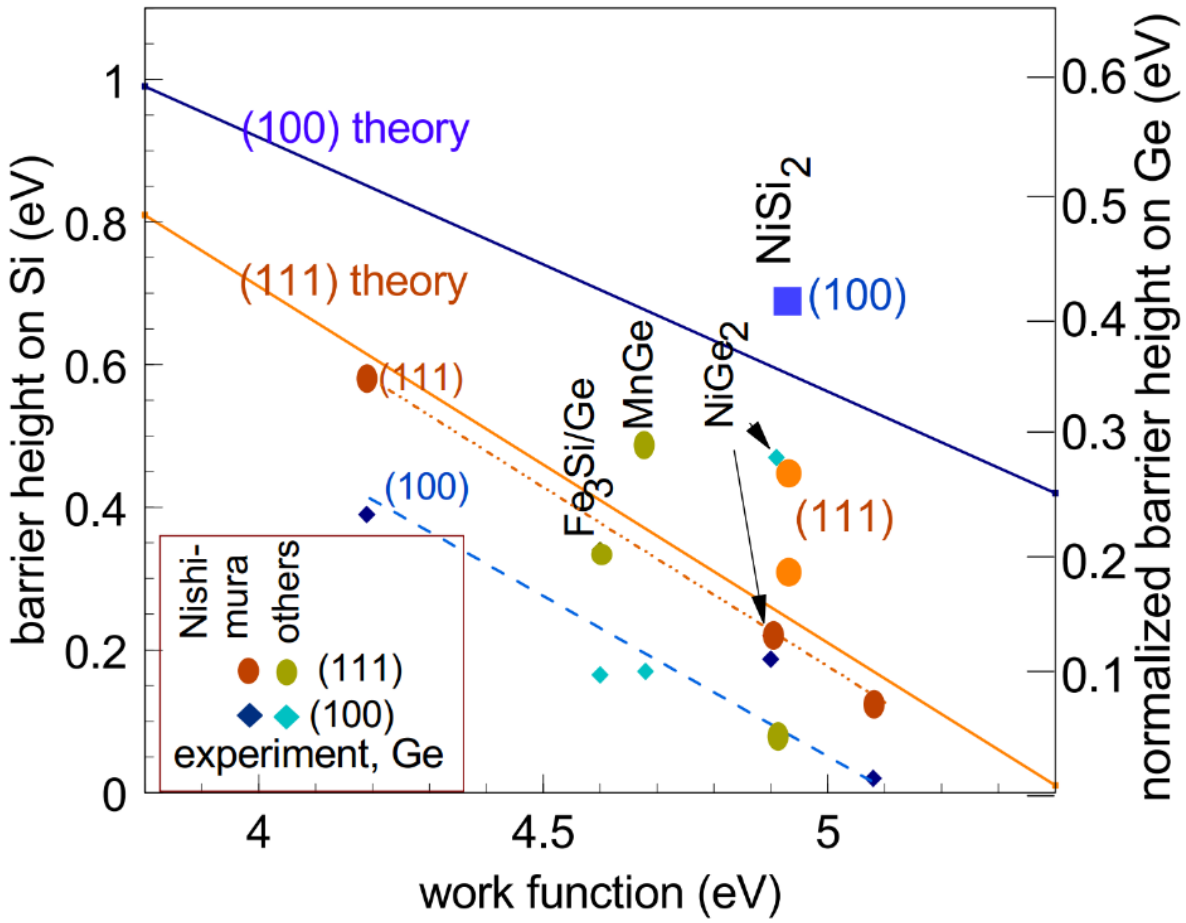
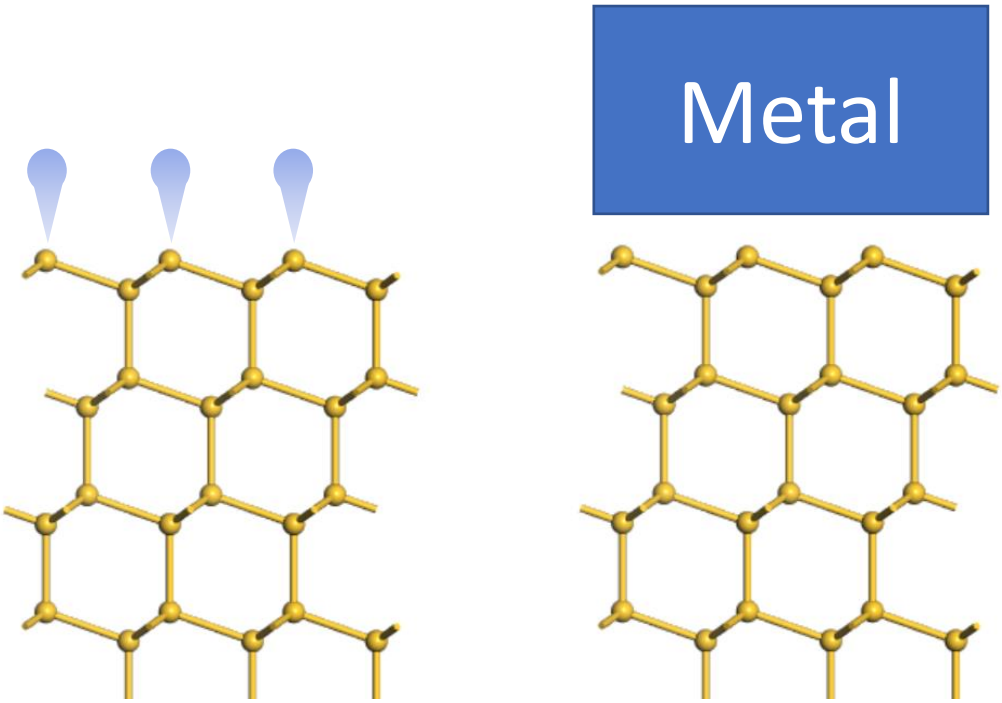
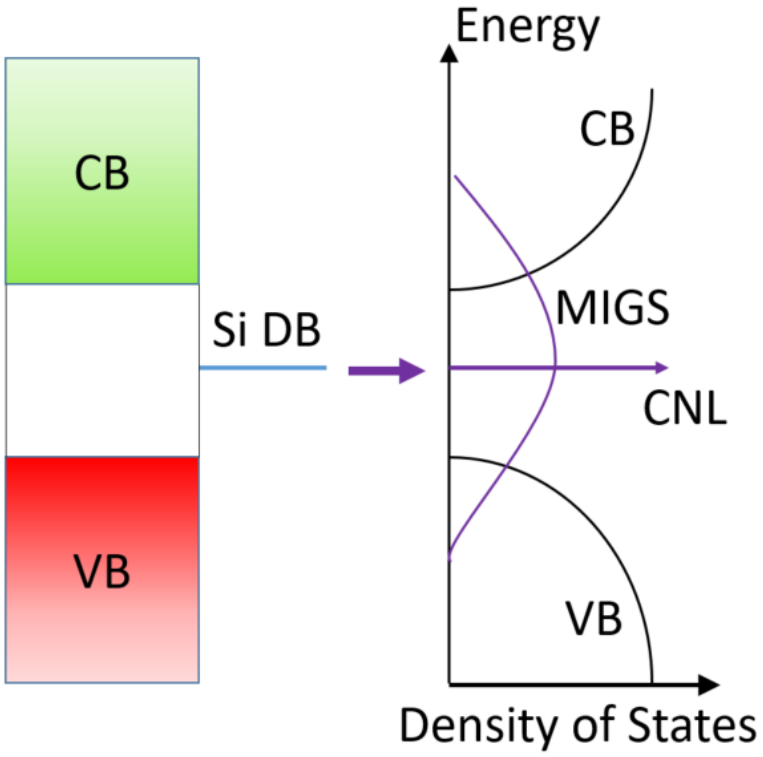


Fig 5.5. Normalised calculated barrierheight values on Si and Ge from **Fig 5.3** and **Fig 5.4**, compared to experimental data for Schottky barrier heights from Refs [12, 22, 33, 35-39]



(a) (b)



(c)

Fig 5.6 (a) Schematic of the Si(111) surface, showing Si dangling bonds, (b) Si/metal interface, (c) schematic band diagram with the Si dangling bond states which have broadened into MIGS.

Our calculated barrier heights for (100) facets lie above those for (111) for both Si and Ge. This is very clear if we normalise the Ge band gap and φ_n values to those of Si, as is done in **Fig 5.5**. Our calculated results agree with experiment for Si/silicides, whose interfaces are abrupt, well-controlled and well characterised. However they are mostly opposite to experiment for Ge, where (111) SBH values often lie above (100) values as seen the experimental points in **Fig 5.5** for Nishimura [22], Yamane [35], Nishimura [36] and Deng [37]. This suggests that Ge/germanide interfaces might be a problem, being disordered, non-abrupt or multi-faceted.

Experimentally it is more difficult to grow epitaxial germanides on Ge [35-39]. This is partly because of missing phases in the phase diagrams of bulk germanides [38] and partly because of poor texture control and micro-crystallinity, which makes their Schottky barriers inhomogeneous and shifts the barrier size. Thus there are fewer well behaved epitaxial Ge/germanide systems for growth [36,37]. Comparing experimental data with our calculated values in **Fig 5.5**, we see that experimentally the (111) SBHs lie above the (100) values for the Ge/$MnGe_x$ system [36], for the range of germanides studied by Nishmura and Toriumi [22] and for Fe_3Si on Ge [35]. In some cases, one facet has an epitaxial layer on it while the other has a polycrystalline layer. On the other hand, there is the useful case of laser-processed $NiGe_2$ where φ_n for the (100) facet is only 0.37 eV, compared to 0.6 eV for NiGe itself [39]. This case is consistent with our calculations. This is a reasonable starting point to create low ϕ_n SBHs for Ge. Thus, theory can help resolve some problems in these new channel materials.

3.6 Explanation to the weak Fermi Level pinning effect of silicides and germanides

What is the cause of the different behavior of elemental metals and silicides? At the Si surface, there are Si dangling bonds. When this Si surface makes an interface with an elemental metal, these dangling bond (DB) states hybridize with the metal states and form the MIGS. These MIGS spread out in energy across the whole bonding-antibonding gap of the Si [15]. The average energy of the MIGS or the CNL energy is therefore that of the Si DB states from which they originated, **Fig 5.6**. Thus, all the properties of these MIGS are intrinsic to the semiconductor.

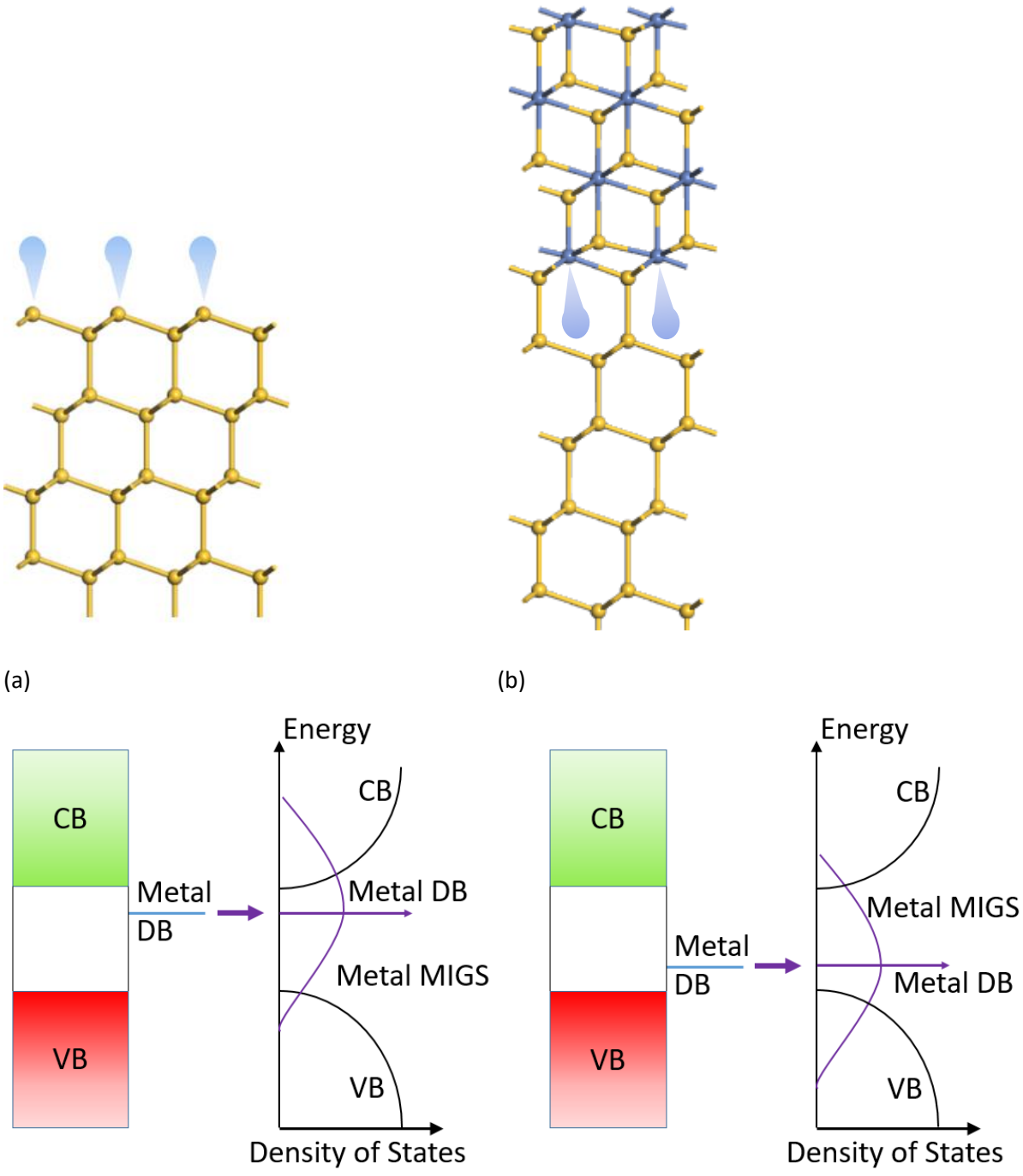


Fig 5.7 (a) Schematic of Si dangling bonds at Si surface, (b) how silicides well passivate these Si dangling bonds at interface, and introduce metal dangling bonds, (c) schematic band diagram show how metal dangling bonds from silicides broaden out to form MIGS at the interface, (d) how MIGS energy varies as metal in silicides change.

Now, at a Si/silicide interface, and taking the Si/$NiSi_2$ interface as an example, the Si-Si bonds continue across the interface. There are no Si DBs on the Si side pointing into the metal

region, the dangling bonds are on Ni atoms on the silicide side. Indeed, the state at E_F for $NiSi_2$(111) interface causing the SBH shift was shown to have Ni DB character by Lin [13]. When the interface forms, these metal DB states spread out in energy across the Ni-Si bonding antibonding energy gap to form the MIGS, **Fig 5.7**. The average energy of these MIGS is now the average energy of the Ni-Si energy gap, or $E_x = ½(E_M + E_{Si})$ in general, where these energies are the work functions of the metal and Si respectively. Hence when the silicide's metal changes, the average energy of the MIGS also changes, by ½ the rate of E_M. The slope factor S to zeroth order will therefore be ~0.5, as indeed it is.

4. Summary

In summary, the barrier heights of a wide range of metal disilicides and metal digermanides have been calculated using supercell models of the interfaces, for the three higher symmetry faces, (100), (111) and (110). The barrier heights show that the SBHs have a dependence of crystal facet, and much weaker Fermi level pinning than for elemental metals, which goes beyond MIGS theory. The dependence on facet is consistent with data for the (111) and (100) interfaces of $NiSi_2$ and can be used to help understand the more complicated behavior of the facet dependence of germanides on Ge. These effects can be used to add lower n-type SBH for Ge, which is presently a significant impediment to the use of Ge high mobility substrates in future CMOS devices.

5. References

[1] Allain, A., Kang, J., Banerjee, K., Kis, A.. Nature Mater. **14**, 1195-1205 (2015)

[2] Dimoulas, A., Tsipas, P., Sotiropoulos, A. & Evangelou, E.K, Appl. Phys. Lett. **89,** 252110 (2006)

[3] Nishimura, T., Kita, K. & Toriumi, A., Appl. Phys. Lett. **91,** 123123 (2007)

[4] Nishimura, T., Kita, K., Toriumi, A., Appl Phys Express **1** 051406 (2008)

[5] K Ikeda, et al. App Phys Lett **88** 152115 (2006)

[6] Freeouf, J. L. Solid State Commun **33,** 1059-1061, (1980)

[7] Tu, K. N., Thompson, R. D. & Tsaur, B. Y. Appl. Phys. Lett. **38,** 626-628 (1981)

[8] Chen, L. J. &Tu, K. N., Mater. Sci. Reports **6,** 53-140 (1991)

[9] Joo, M. S.,et al. J. Vac. Sci. Technol. B **24,** 1341-1343 (2006)

[10] Bucher, E. & Schulz, S., et al. Appl. Phys. A **40,** 71 (1986)

[11] Vandre, S., Kalka, T., Preinesberger, C. & Dahne-Prietsch, Phys. Rev. Lett. **82** 1927-1930 (1999)

[12] Tung, R. T., App. Phys. Rev. **1,** 011304 (2014)

[13] Lin, L., Guo, Y. & Robertson, J. Appl. Phys. Lett. **101**, 052110 (2012)

[14] Tersoff, J. Phys. Rev. Lett. **52,** 465-468 (1984)

[15] Monch, W., Phys. Rev. Lett. **58,** 1260-1263 (1986)

[16] Robertson, J., Phys. Stat. Solidi. A **207,** 261-269 (2010)

[17] Robertson, J.J. Vac. Sci. Technol. B **18**, 1785-1791 (2000)

[18] Spicer, W. E., Lindau, I., Skeath, P., Su, C. Y. & Chye, P. Phys. Rev. Lett. **44,** 420-423 (1980)

[19] Das, S., Chen, H. Y., Penumatcha, A. V., Apenzeller, J., Nano Letts **13**, 100-105 (2013)

[20] Guo, Y., Liu, D., Robertson, J., ACS Appl Mater Interfaces **7**, 25709-25715 (2015)

[21] Liu, D., Guo, Y., Fang, L., Robertson, J., App Phys Lett **103**, 183113 (2013)

[22] Nishimura, T., Yajima, T. & Toriumi, A, Appl. Phys. Exp. **9** 081201 (2016).

[23] Tung, R. T., Phys. Rev. Lett. **52**, 461-464 (1984)

[24] Clark, S. J., et. al., Z. Kristall. **220,** 567 (2005)

[25] Clark, S. J. & Robertson, J., Phys. Rev. B **82**, 085208 (2010)

[26] Das, G. P, et al, Phys Rev Lett **63,** 1168 (1989)

[27] Cherns, D., Anstis, D. R., Hutchinson, J.H. & Spence, J. C. H.. Philos. Mag. A**46,** 849 (1982)

[28] Cherns, D., Hetherington, C. J. D. & Humphreys, C. J., Philos. Mag. A **4B**, 165 (1984)

[29] Loretto, D., Gibson, J. M., & Yalisove, J. M., Phys. Rev. Lett. **63**, 298-301 (1989)

[30] Yu, B. D., et al., J. Vac. Sci. Technol. B **19**, 1180 (2001)

[31] Falke, U., Bleloch, A., Falke, M., & Teichert, S., Phys. Rev. Lett. **92,** 116103 (2004)

[32] Michaelson, M. B. J. Appl. Phys. **48** 4729 (1977)

[33] Tung, R. T., J. Vac. Sci. Technol. B **11**, 1546 (1993)

[34] Tung, R. T., et al., Phys. Rev. Lett. **66**, 72-75 (1991)

[35] Yamane, K., et al., Appl. Phys. Lett. **96,** 162104 (2010)

[36] Nishimura, T., et al., Microelec. Eng. **88**, 605 (2011)

[37] Deng, Y., et al., Thin Solid Films **557,** 84-89 (2014)

[38] DeSchutter, B., DeKeyser, K., Lavoie, C., & Detavernier, C. Appl. Phys. Rev. **3,** 031302 (2016)

[39] Lim, P. S. H., Chi, D. Z., Wang, X. C. & Yeo, Y. C., Appl. Phys. Lett. **101**, 172103 (2012)

Chapter 6. Oxygen vacancy in TiO_2 for Application in ReRAM

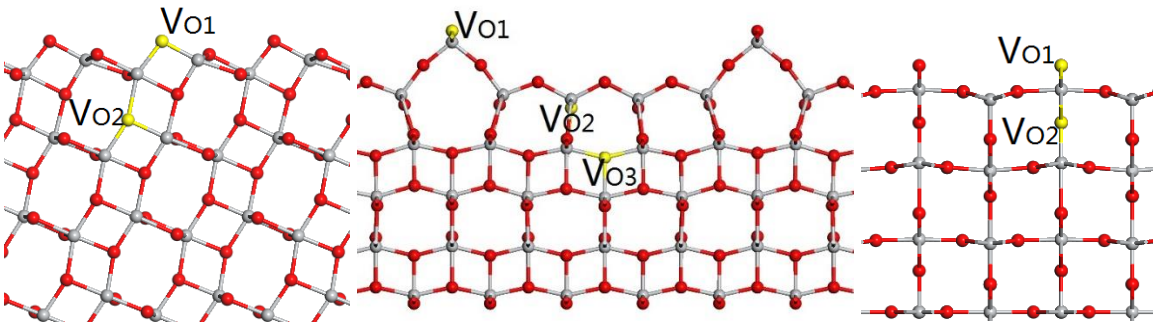
1. Overview

Titanium oxide (TiO_2) is a technologically important transition metal oxide [1] widely used in photocatalysis [2], solar energy conversion [3], environmental cleanup [4] and may be in future resistive random-access memories (RRAM) [4]. The oxygen vacancy (V_O) is the most important point defect in TiO_2 and it plays a crucial role in these applications. Electron bombardment results confirm that oxygen vacancies rather than Ti interstitials make the dominant contribution to the band gap states of TiO_2 surface [5]. Oxygen vacancies also influence the absorption of water and other molecules on TiO_2 surfaces in photocatalytic devices [6]. Similarly, oxygen vacancies form a conductive filament across the TiO_2 film in RRAM devices [4,7]. In addition, the oxygen vacancy usually forms at the surface region of TiO_2 in experiments so the surface defects are more important than bulk ones [8,9]. Therefore, it is of great significance to understand the behavior of oxygen vacancies at TiO_2 surfaces and associate it with the performance of TiO_2 surfaces and devices.

While rutile is the thermodynamically most stable bulk phase of TiO_2, anatase is more stable for particles below ~14 nm in size due to its lower surface energies [10], and so anatase is more relevant to catalytic applications. There has been a continuing interest in the properties of anatase and rutile surfaces [1]. Rutile (110), its most stable surface, has a high surface energy ($0.82 J/m^2$), whereas anatase has a much lower surface energy of $0.44 J/m^2$ for its most stable (101) facet. The more reactive (001) facet of anatase is calculated to have a surface energy of ~$0.90 J/m^2$. However, this surface can undergo a surface reconstruction to give the ad-molecule (ADM) reconstruction, as confirmed by STM [11], with a calculated surface energy of 0.51 J/m^2. This minority TiO_2 facet has aroused much attention since it was fabricated by stabilizing with a hydrofluoric acid morphology controlling agent [12].

As O vacancies mediate the reactivity of TiO_2 surfaces, it is expected that the surface with the higher vacancy concentration might be the most reactive. However, STM experiments suggest that anatase has a relatively low vacancy concentration [13]. Cheng and Selloni [14] explained this by showing that the O vacancy was more stable subsurface in anatase, whereas it was more stable on the surface in rutile. Nevertheless, these results were obtained by the simple generalized gradient approximation (GGA) version of density functional theory. It is well known that that DFT and GGA under-estimates in semiconductors and insulators [15]. In TiO_2 this has the additional effect that it under-estimates the wavefunction localization and degree

of lattice distortions at vacancies, which lead to them being deep defects [16,17]. Thus, it is unclear if predictions of vacancy behavior between the bulk and surfaces in TiO_2 based only on GGA would be fully reliable. It is therefore worthwhile to check these results using hybrid functional methods which correct these errors. The simplest method to use would be DFT+U which includes an on-site repulsion potential U [18]. However, this requires rather physical values of U in the case of TiO_2. In addition, the on-site repulsion U only opens the band gap but doesn't give correct description of band shape, which is detrimental to electronic properties of TiO_2. Thus, it is preferable to use hybrid functional such as the Heyd, Scuseria, Erzernhof (HSE) method or screened exchange (sX) methods [19] which give good representations of the electronic structure, while being sufficiently computationally efficient that they can be used for large supercells of 100-200 atoms.



(a) (b) (c)

Fig 6.1. Slab models of (a) anatase (101) surface, (b) anatase (001) surface and (c) rutile (110) surface. Oxygen is red, titanium is grey. The various oxygen vacancies are yellow.

We have previously presented the results for oxygen vacancies in bulk rutile by sX [20]. Similar results using HSE and GW methods are also available [21-23]. Here we present the results for the surface oxygen vacancies on TiO_2 by sX. To this end, we calculate the formation energy of oxygen vacancies at three surfaces, anatase (101), reconstructed anatase (001), and rutile (110) by the sX hybrid functional to find out the relative stability of the surface and subsurface positions. The sX hybrid functional method was found to give the accurate band gaps of many semiconductors and insulators, and the correct localization of charge near vacancy [24,25]. In this paper we include the results of GGA [26] as well for comparison.

2. Computational Setups

We carried out our GGA and screened exchange hybrid functional simulation using the CASTEP plane wave pseudopotential code [27]. For k sampling we took the Γ point scheme due to the large supercell. The geometry optimization was carried out until residual forces on each atom was smaller than 0.01eV/Å.

The three TiO_2 surfaces we constructed are shown in **Fig 6.1**. We used slabs of four TiO_2 layers containing 144 atoms for the anatase (101) surface, and slabs of four TiO_2 layers containing 96 atoms for rutile (110) surface. For the anatase (001) surface, we took the ADM model [11] as introduced above, with five TiO_2 layers containing 189 atoms. All surfaces were modeled using the periodical boundary condition with a vacuum of ~ 11Å.

The defect properties are represented by their formation energies as a function of Fermi energy and chemical potential. The formation energy H_q of defect with charge q can be calculated by following expression [28]

$$H_q(E_F,\mu) = [E_q - E_H] + q(E_v + \Delta E_F) + \sum_\alpha n_\alpha(\mu_\alpha) \quad (1)$$

where E_q and E_H are the total energy of a defect cell and a perfect cell with charge q respectively, ΔE_F is the Fermi level energy with respect to valence band maximum, n_α is the number of atoms of element α, and μ_α is reference chemical potential of element α. The chemical potentials satisfy the following equation:

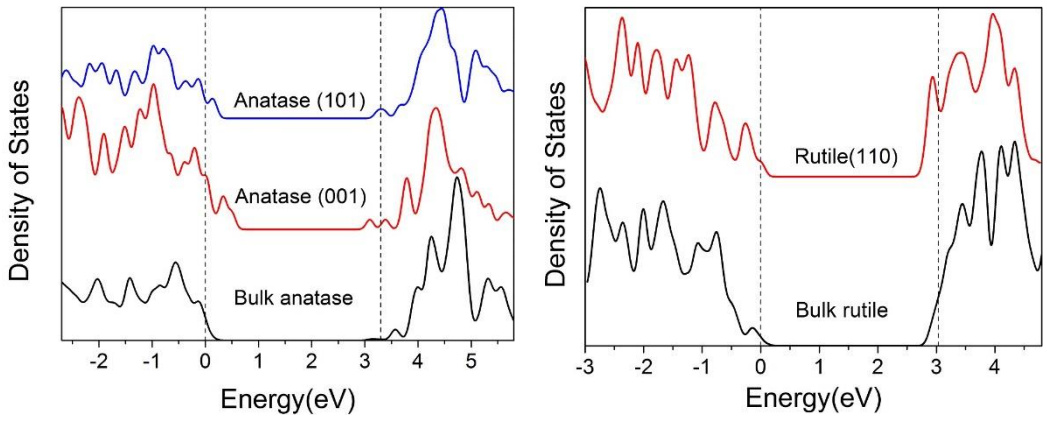
$$\mu_{Ti} + 2\mu_O = H_f(TiO_2) = -9.73 \text{ eV} \quad (2)$$

The O-rich limit is at $\mu_O = 0$ eV, and $\mu_{Ti} = -9.73$ eV, while the O-poor limit (or Ti-rich limit) is at $\mu_O = -4.07$ eV and $\mu_{Ti} = -1.59$ eV which corresponds to the equilibrium of TiO_2 and Ti_2O_3.

3. Results

The density of states (DOS) of the three TiO_2 surfaces are shown in the **Fig 6.2**, and compared to the bulk TiO_2 DOS. The calculated bulk TiO_2 band gaps are 3.02eV for rutile and 3.17eV for anatase, in close to the experimental value. The dangling bonds on the surface introduced the gap states expanding into the band gap of bulk TiO_2. The surface band gaps are thus smaller than the bulk value, 3.11eV for anatase (101), 2.76eV for anatase (001) and 2.90eV for rutile

(110). Our simulation results are consistent with the experimental surface band gap values in ref [29].



(a) (b)

Fig 6.2 Densities of states of (a) anatase (101) surface, anatase (001) surface and bulk anatase; (b) rutile (110) surface and bulk rutile, calculated by sX-LDA method. The dashed line marks the conduction band minimum and valance band maximum of bulk TiO_2

We have calculated the formation energy in the O-rich condition of inequivalent surface or subsurface oxygen vacancies on all three surfaces by GGA and screened exchange, and the most stable ones are listed in **Table 6.1**. The formation energy of oxygen vacancy in bulk TiO_2 is also listed here for comparison.

Table 6.1. Formation energies of oxygen vacancy defect at different sites of anatase (101), anatase (001), and rutile (110) surfaces.

Formation Energy	Vacancy Site	GGA(eV)	sX-LDA(eV)
anatase (101)	V_{O1}	4.37	5.08
	V_{O2}	**3.99**	**4.78**
	V_O in bulk	4.27	4.95
anatase (001)	V_{O1}	4.74	5.07
	V_{O2}	4.38	5.12
	V_{O3}	**4.21**	**4.96**
	V_O in bulk	4.27	4.95
rutile (110)	V_{O1}	**3.71**	**4.39**
	V_{O2}	4.10	4.98
	V_O in bulk	4.79	5.70

The bold denotes the most stable defect site for each surface. The values are compared with oxygen vacancy in bulk TiO_2. Site labels are in Fig. 6.1.

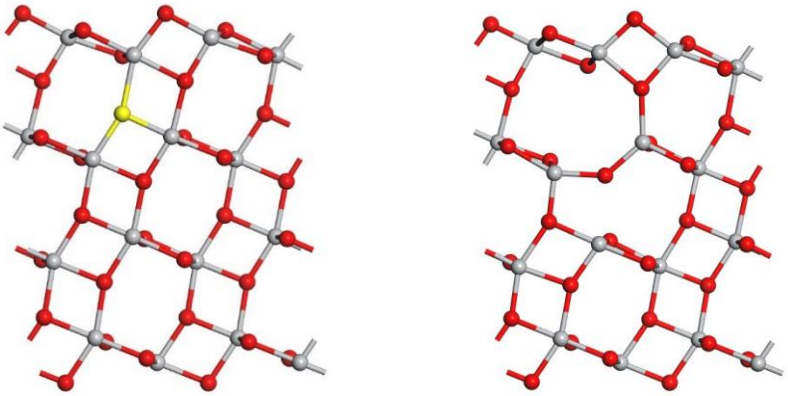
First we discuss the most stable oxygen vacancies at anatase (101) surface. The surface defect site V_{O1} corresponds to the removal of an O_{2C}, while V_{O2}, the subsurface defect site, corresponds to the removal of an O_{3C}. Both GGA and sX find that V_{O2} is more favorable, with an energy difference of 0.3eV. This result is consistent with the quite low surface energy of anatase (101), thus it takes more energy to create a *surface* vacancy. In addition the subsurface oxygen vacancy formation energy is also smaller than that in bulk anatase, so the oxygen vacancy is more likely to form at anatase (101) subsurface rather than the surface or bulk site.

We then analyze the three most stable oxygen vacancies on anatase (001). V_{O1} is at the ridge part of the surface, which is originally bonded to two Ti atoms. V_{O2}, corresponding to the removal of an O_{3C}, is at the terrace part of the surface, and V_{O3} is at the subsurface which is originally bonded to three Ti atoms. Both GGA and sX show that the subsurface oxygen vacancy V_{O3} is most stable, similar to anatase (101). In addition, the formation energy of oxygen vacancy in bulk anatase is 4.95eV, which is nearly the same with that of V_{O3} in anatase (001) surface, while formation energy of V_{O2} in anatase (101) is 0.2eV smaller. Thus, the subsurface oxygen vacancy in anatase (001) experiences an environment similar to the bulk TiO_2 and the oxygen vacancy is likely to form at subsurface and bulk site rather than surface. Furthermore, we can also conclude that O vacancies prefer to form at the subsurface rather than surface of anatase (101) and (001).

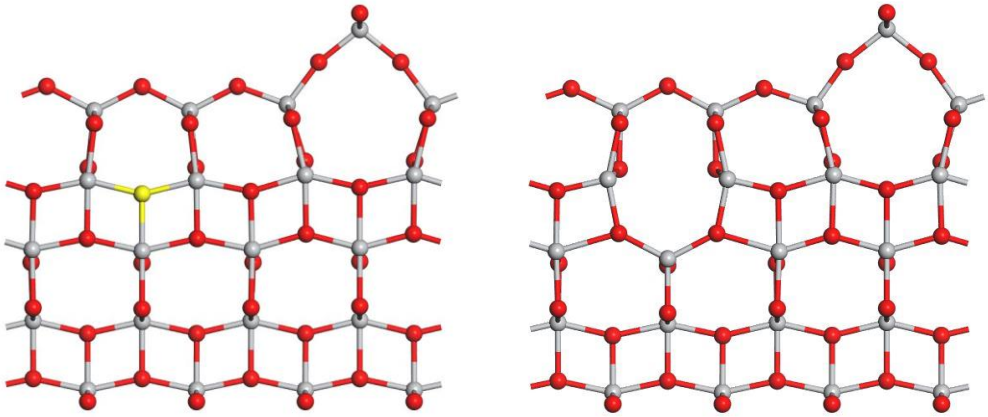
For rutile (110), **Table 6.1** lists two most stable oxygen vacancies that we get in our calculation. V_{O1} is at the surface, which is originally bonded to two Ti atoms while V_{O2} is a subsurface defect which corresponds to the removal of an O_{3C} atom. Both GGA and sX show that V_{O1} is the most stable, with a formation energy 1.3eV lower than V_O in bulk rutile. Thus, oxygen vacancies are more likely to form at the rutile surface. This result could be interpreted as that

it only need to break two Ti-O bonds to create an oxygen vacancy at the rutile (110) surface rather than to break three Ti-O bonds at subsurface or bulk.

(a)



(b)



(c)

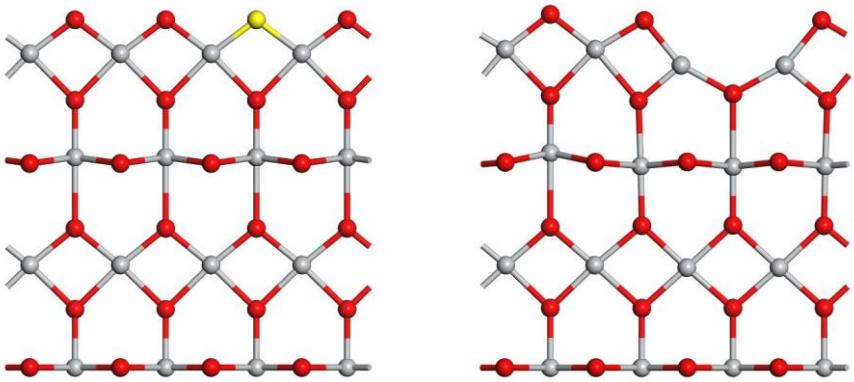
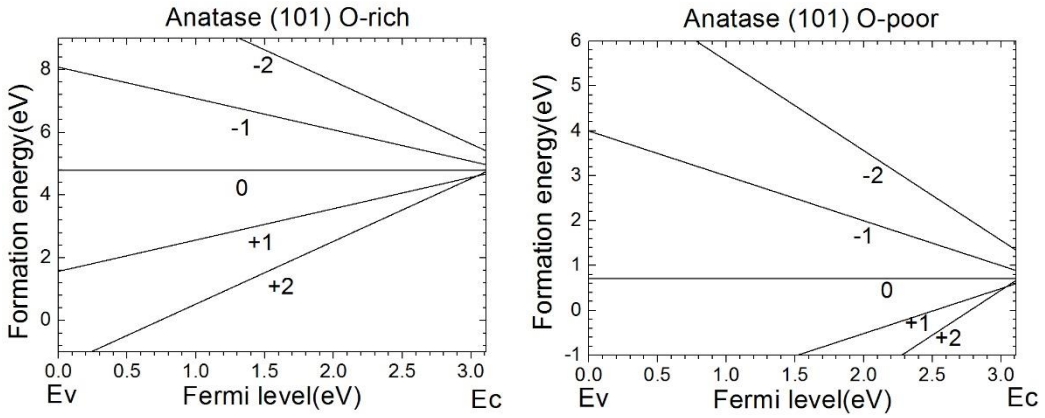


Fig 6.3. Side view of atomic spacing of (a) anatase (101) surface, (b) anatase (001) surface and (c) rutile (110) surface. The figures on the left are perfect slabs without defect, and the figures on the right are the surface structure with an oxygen vacancy after geometry relaxation.

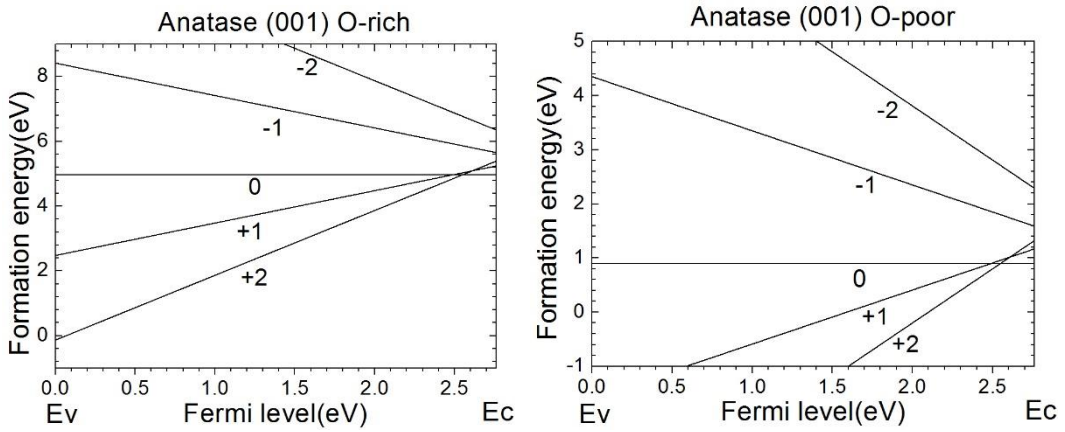
Removal of an oxygen atom in TiO_2 surface structure creates two unpaired electrons (neutral defect state) and dangling bonds on adjacent Ti atoms. For the anatase (101) surface, removal of oxygen atom at V_{O2} site shifts the adjacent Ti atoms away from defect, and nearby oxygen atoms towards defect. The Ti-Ti distance increase from 3.20Å to 3.70Å and from 4.00Å to 4.75Å, and a Ti-O bond near defect breaks due to the distortion. For anatase (001) surface, the removal of oxygen atom at V_{O3} site also causes adjacent Ti atoms to move apart and O atoms to move close to defect. The Ti-Ti distance between adjacent Ti atoms increase from 3.77Å to 4.12Å and from 3.08Å to 3.43Å, and the O-O distance between nearby O atoms decrease from 3.74Å to 3.34Å. For the rutile (110) surface, the most stable oxygen vacancy site is at the surface, so the removal of oxygen atom only forms two Ti dangling bond. It not only causes the adjacent Ti atoms to move away from defect, increases the distance from 2.96Å to 3.37Å, but also triggers these two Ti atoms to move 0.40Å downwards into bulk. However, the positions of the next-nearest Ti atoms stay nearly unchanged, thus the oxygen vacancy only lead to a local deformation in TiO_2.

Fig 6.4 shows the calculated formation energies of most stable oxygen vacancies at these three surfaces for both O-poor and O-rich conditions by sX, as a function of Fermi level. Defect charge states including -2, -1, 0, 1, 2 have been calculated and spin polarization is introduced in our simulation. The E_C and E_V here refer to the conduction band and valence band edge of the TiO_2 surface system, so the band gap is smaller than bulk value. The charge transition levels correspond to the Fermi level at which the charge q and q' defect states have the same formation energy, so the defect charge state would change at these energies.

(a)



(b)



(c)

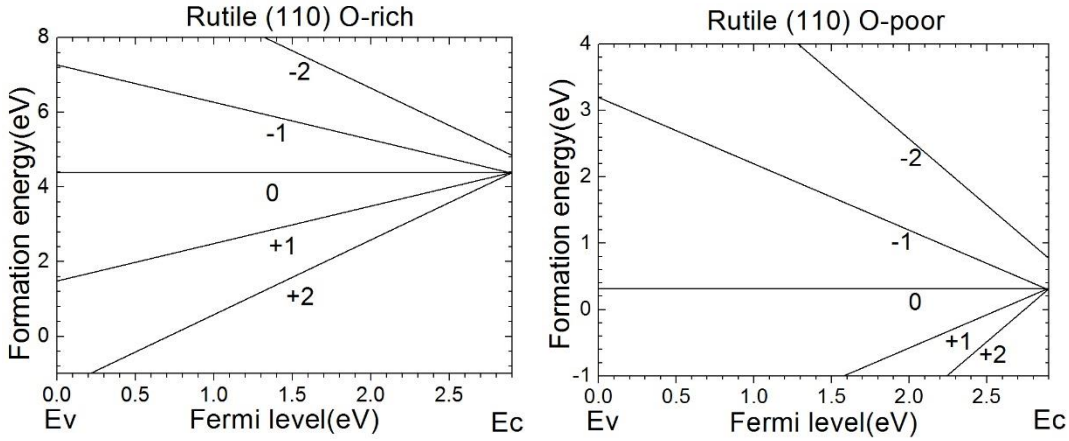


Fig 6.4 The oxygen vacancy formation energy against Fermi level at O-rich and O-poor conditions, calculated by sX-LDA for (a) anatase (101), (b) anatase (001), and (c) rutile (110) surface. E_C and E_V refer to the conduction band and valance band edge of TiO_2 surface.

On the anatase (101) surface, the most stable oxygen vacancy V_{O2} is mostly dominated by the +2 charge state, and the calculated V_O^+/V_O^{2+} charge transition is just 0.1eV below the conduction band minium (CBM). At O-rich limit, the formation energy of V_O^{2+} charge state is -1.48eV at the valence band maximum (VBM), and this value increases to 4.60eV near conduction band. Thus, oxygen vacancy could form spontaneously at O-rich condition when Fermi level is near valence band. On the contrary, oxygen vacancies could form spontaneous regardless of the position of Fermi level at O-poor condition, such as the metal-insulator interface in RRAM.

On the anatase (001) surface, the formation energy of V_{O3} is -0.13eV at VBM and around 4.96eV at CBM, at the O-rich condition. The V_O^0/V_O^{2+} charge transition is at 0.2eV below the conduction band. The +1 charge state is unstable, so it is a negative U system [30]. For the O-poor condition, formation energy of V_{O3} is negative in most part of band gap, therefore oxygen vacancy could form spontaneously.

As for the rutile (110) surface, the most stable surface defect V_{O1} has the formation energy of -1.42eV near VBM at O-rich limit, thus the oxygen vacancy could form spontaneously on this surface when Fermi level is near VBM. The V_O^{2+} charge state has the lowest formation energy throughout the whole band gap, so in the device involving rutile (101) surface, Fermi level could move freely without the charging process of oxygen vacancies. In addition, at O-poor limit, the formation energy of V_{O1} is negative in most region of band gap.

Fig 6.5 shows the partial density of states (PDOS) of the most stable defect sites for charge state V_O^{2+} on three surfaces. Here we compared the total DOS, the PDOS on Ti atoms near defect and in the bulk, as well as the PDOS on O atoms in bulk. In sX, the oxygen vacancy gives rise to a defect state just above CBM on anatase (101) and anatase (001) surface, while a defect state appears below CBM on rutile (110) surface.

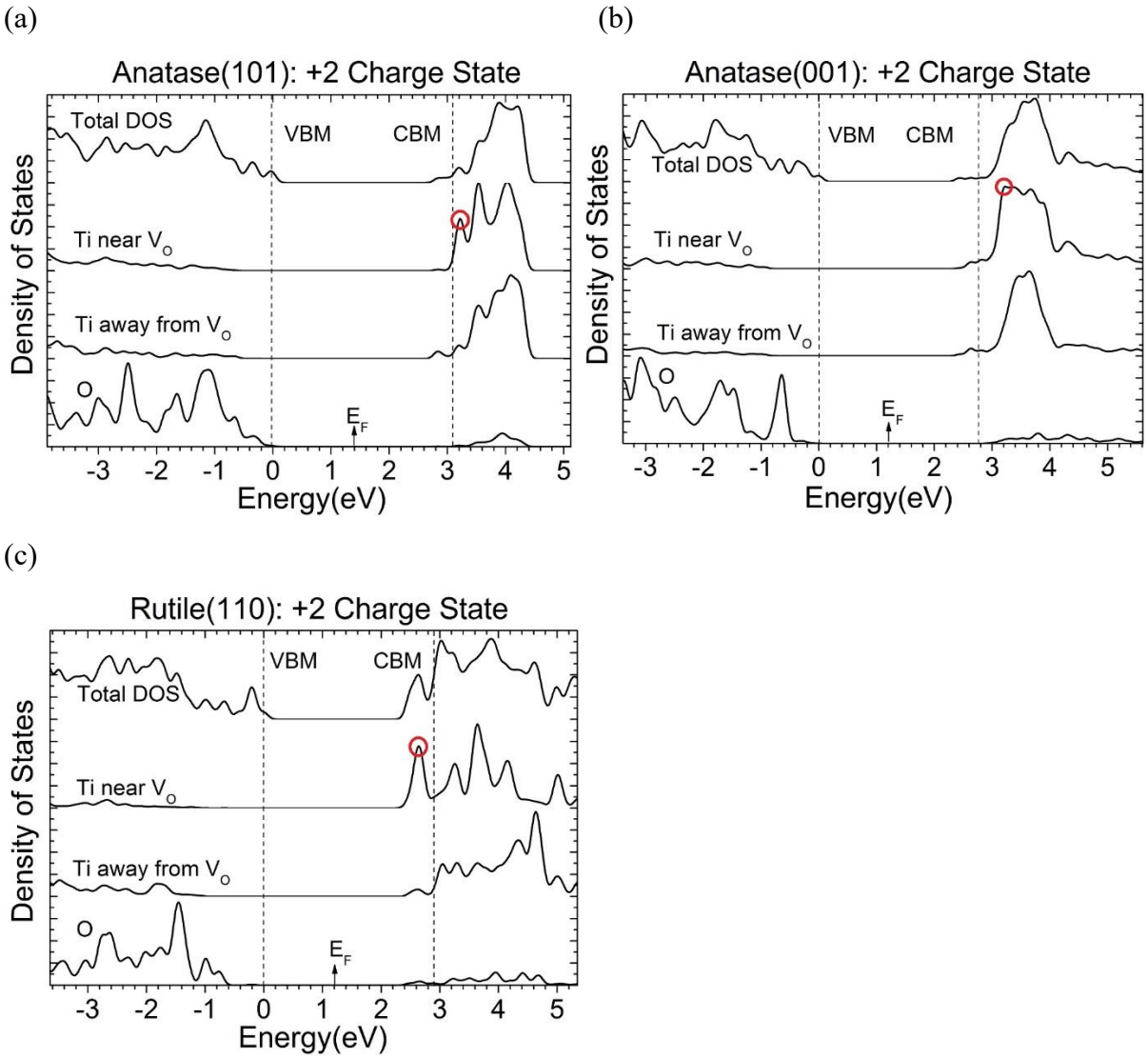
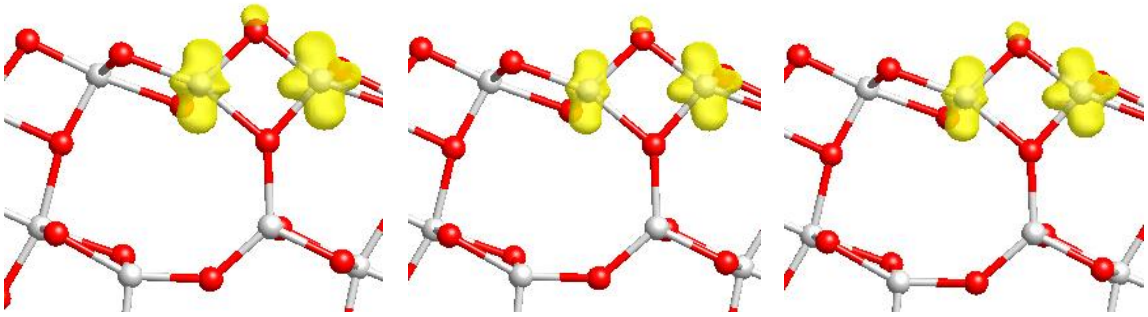


Fig 6.5 The partial density of states (PDOS) of the three TiO_2 surfaces with an oxygen vacancy at +2 charge states, (a) anatase (101) surface, (b) anatase (001) surface, and (c) rutile (110) surface, by sX-LDA method. The PDOS on Ti atom near the defect, on Ti atom away from defect and on O atom are shown here and compared with total DOS. The VBM and CBM are calculated from the perfect surface slab model and aligned to the PDOS of defect system. The defect peak is indicated with right circle in each figure.

Fig 6.6 shows the charge density of defect states for V_{O2}^{0}, V_{O2}^{+}, and V_{O2}^{2+} on anatase (101) viewed from same direction. Different charge state of the defect triggers the different lattice distortion in neighbouring area. Interestingly, the orbital is localized on two Ti atoms rather than the defect site. One of the Ti atom is the nearest neighbour and the other is the next-nearest

neighbour, while both of them are at the surface. This situation is different from that in bulk TiO_2 as described in ref [20], maybe due to the influence of surface. However, this localization scheme is consistent with the B3LYP and PBE+U results by Pacchioni et al [17]. We have also calculated the charge density of defect level on anatase (001) and rutile (110) surface, both of them are delocalized and not shown here.



(a) (b) (c)

Fig 6.6. The charge density contour for oxygen vacancy site of anatase (101) surface: (a) V_{O2}^{0}, (b) V_{O2}^{+}, (c) V_{O2}^{2+}

4. Discussions

Various experimental studies have confirmed the existence of oxygen vacancies on the surface of rutile (110), which are in good agreement with our simulation results. K. Fukui et al. directly observed the oxygen vacancies on rutile (110) surface via AFM in 1997 [8]. S. Wendt et al. also confirmed the existence of oxygen vacancies on this surface by STM in 2005 [9]. Both AFM and STM figures of surface oxygen vacancies provide strong supports to our simulation results that oxygen vacancies are more likely to form on the surface of rutile (110), since the formation energy of oxygen vacancy at rutile (110) surface is smaller than that at subsurface and much smaller than that in bulk rutile.

We compare our sX simulation results of neutral V_O formation energy with the GGA results by Cheng et al. [14]. Both GGA and sX show that subsurface Vo is most stable for the two anatase surfaces while surface Vo is most stable for rutile (110) surface. However, our charge transition level and PDOS results suggest that it is the +2 charge state rather than the neutral state that has the lowest formation energy in most part of band gap for all three surfaces. The oxygen vacancies with +2 charge can form spontaneously at O-poor limit regardless of the position of Fermi level, and the Fermi level could move freely without pinning.

We also take larger supercells for all three surface models than in Cheng et al. [14]. In our simulation, a distance of at least 10Å is ensured between defect and its image in all directions, and the slab model is thick enough along z direction to limit the influence of bottom surface. Our cell size test indicates that the small supercell could lead incorrect formation energy order. For the anatase (001) surface, the four slabs model we have tested is not thick enough that the subsurface V_{O3} could be influenced by both the top surface and the bottom surface in the model, which results in an unreasonable charge distribution and formation energy.

We also compare our results of Vo on surface and in bulk with other advanced methods beyond GGA such as GW. It is widely agreed that neutral Vo formation energy in bulk rutile at O-poor limit is around 1.8eV, which is also confirmed by our simulation. However, the +2 charge state of Vo formation energy differs largely in different work. HSE result given by Janotti et al [21] suggests this value to be -5.1eV while GW result of Malashevich et al [22] gives the formation energy of -3.7eV. Our sX results give -4.30eV, which is in the reasonable region. By comparing the oxygen vacancy formation energy on surface and in bulk, we find that anatase has a similar formation of Vo on surface as it is in bulk, while rutile Vo on surface is thermodynamically much preferred than in bulk Vo. Thus, V_O is likely to form in the surface area, while for anatase surface the oxygen vacancy tends to form at the subsurface and bulk region.

Our simulation also shows that formation energy of oxygen vacancy at anatase (101) subsurface is smaller than that at anatase (001). On the other hand, the formation energy of anatase (001) subsurface is nearly the same with that in bulk anatase, so the surface and subsurface region does not possess any advantage over bulk for oxygen vacancy's formation. These two mechanisms jointly explain the reason for a much lower oxygen vacancy concentration on anatase (001) surface than anatase (101) surface as indicated by resonant photoemission data [13].

5. Conclusions

We calculate the electronic structure of oxygen vacancies with different charge states on three TiO_2 by screened exchange functional. On all of the three surfaces, the +2 charge state is most stable for oxygen vacancy in most part of band gap. The formation energy of Vo is negative near VBM at O-rich limit so that O vacancy could form spontaneously at low Fermi level or O poor condition. This is favorable to the formation of the conducting channel in RRAM. The subsurface oxygen vacancy is most stable on anatase (101) and anatase (001) surface, while these defects are more likely to form at the surface of rutile (110) rather than the subsurface or bulk sites. For all three TiO_2 surface models, oxygen vacancies introduce a gap state near the conduction band edge. The defect orbital is localized on a neighboring Ti atom and a next-nearest Ti atom, which is different from bulk TiO_2 defect.

6. References

[1] Diebold, U. Surf. Sci. Rep. 2003, 48, 53-229.

[2] Asahi, R.; Morikawa, T.; Ohwaki, T.; Aoki, K.; Taga, Y. Science 2001, 293, 269-271.

[3] M. Gratzel. Nature 2001, 414, 338-344.

[4] Kwon, D. H. et al, Nature Nanotech 2010, 5, 148-153.

[5] Yim, C. M.; Pang, C. L.; Thornton, G. Phys. Rev. Lett. 2010, 104, 036806.

[6] Schaub, R.; Thostrup, P.; Lopez, N.; Laegsgaard, E.; Stensgaard, I.; Norskov, J. K.; Besenbacher, F. Phys. Rev. Lett. 2001, 87, 266104.

[7] Park, S. G.; Magyari-Kope, B.; Nishi, Y. IEEE ED Lett. 2011, 32, 197-199.

[8] Fukui, K.; Onishi, H.; Iwasawa, Y. Phys. Rev. Lett. 1997, 79, 4202-4205

[9] Wendt, S. et al. Surf. Sci. 2005, 598, 226-245.

[10] Ranade, M. R. et al. Proc. Natl. Acad. Sci. USA 2002, 99, 6476-6481.

[11] Lazzeri, M.; Selloni, A. Phys. Rev. Lett. 2001, 87, 266105.

[12] Yang, H. G.; et al., Nature 2008, 453, 638.

[13] Thomas, A. G.; et al. Phys. Rev. B. 2007, 75, 035105.

[14] Cheng, H. Z.; Selloni, Phys. Rev. B. 2009, 79, 092101.

[15] Mori-Sanchez, P.; Cohen, A. J.; Yang, W. T. Phys. Rev. Lett. 2008, 100, 146401; G. Pacchioni, G. J. Chem. Phys. 2008, 128, 182505.

[16] Di Valentin, C.; Pacchioni, G.; Selloni, A. Phys. Rev. Lett. (2006) 97 166803.

[17] Di Valentin, C.; Pacchioni, G.; Selloni, J. Phys. Chem. C 2009, 113, 20543-20552.

[18] B. J. Morgan, G. W. Watson. Surface Sci. (2007) 601 5034.

[19] Bylander, D. M.; Kleinman, L. Phys. Rev. B 1990, 41, 7868-7871; Seidl, A.; Gorling, A.; Vogl, P.; Majewski, J. A.; Levy, M. Phys. Rev. B 1996, 53, 3764-3774.

[20] Lee, H. Y.; Clark, S. J.; Robertson, J. Phys. Rev. B 2012, 86, 075209.

[21] Janotti, A.; Varley, J. B.; Rinke, P.; Umezawa, N.; Kresse, G.; Van de Walle, C. G. Phys. Rev. B 2010, 81, 085212.

[22] Malashevich, A.; Jain, M.; Louie, S. G. Phys. Rev. B 2014, 89, 075205.

[23] Zhanpeisov, N. U.; Fukumura, H. J. Phys. Chem. C 2007, 111, 16941-16945; Ortega, Y.; Hevia, D. F.; Oviedo, J.; San-Miguel, M. A. Appl. Surf. Sci. 2014, 294, 42-48; Zhu, L. G.; Hu, Q. M.; Yang, R. J. Phys-Condens Mat. 2014, 26, 055602.

[24] Xiong, K.; Robertson, J.; Gibson, M. C.; Clark, S. J. Appl. Phys. Lett. 2005, 87, 183505; Clark, S. J.; Robertson. J. Phys. Rev. B 2010, 82, 085208

[25] Guo, Y. Z.; Clark, S. J.; Robertson, J., J. Phys-Condens Mat. 2012, 24, 325504.

[26] Perdew, J. P.; Burke, K.; Ernzerhof, M., Phys. Rev. Lett. 1996, 77, 3865-3868.

[27] Clark, S. J.; Segall, M. D.; Pickard, C. J.; Hasnip, P. J.; Probert, M. J.; Refson, K.; Payne, M. C., Z. Kristallogr 2005, 220, 567-570.

[28] Lany, S.; Zunger, A., Phys. Rev. B 2008, 78, 235104.

[29] Goniakowski, J.; Noguera, C., Surf. Sci. 1994, 319, 68-80; Majumder, S.; Paramanik, D.; Solanki, V.; Bag, B. P.; Varma, S., Appl. Phys. Lett. 2011, 98, 053105.

[30] Baraff, G. A.; Kane, E. O.; Schluter, M., Phys. Rev. B 1980, 21, 5662-5686.

Chapter 7. Hydrogen in ZnO and NBIS

1. Overview

Amorphous oxide semiconductors (AOS) such as In,Ga,Zn oxide (IGZO) are now widely used as thin film transistors (TFTs) in high resolution liquid crystal displays and as driver transistors of organic light emitting diode (OLED) displays, because of their higher electron mobility than hydrogenated amorphous silicon (a-Si:H) [1-5]. However, there remains the problem that the TFTs suffer from the negative bias light induced stress instability (NBIS), a shift of the gate threshold voltage during negative bias stress and illumination [6-11]. In NBIS, a slightly sub-bandgap light excites electrons from some filled defect states just above the valence band edge to the conduction band, where they produce a persistent photoconductivity (PPC) [7,11]. The defect states responsible are seen in hard X-ray photo-emission[12], but their cause is still contentious.

There have been various models to account for this instability, involving oxygen vacancies [11,13-16], oxygen interstitials [17-20], or hydrogen impurities [18,21,22], but each of these models has its flaws. The instability has often been attributed to the oxygen vacancy as this is a common defect in oxides and the instability is worse in O-deficient films [15,16]. The O vacancy also has the 'negative-U' property [23,24] which is a mechanism that creates an energy barrier to carrier recombination and thus it explains the persistence of the photoconductivity[11]. However, the problem is that the vacancy states lie too high in energy in the ZnO band gap [24,25] and IGZO (unpublished, in preparation) compared to their energy as seen by photoemission[12].

The second model based on oxygen interstitials does produce states in the required energy range [17,18], as seen in random network models[19], and these can also have the negative-U like property [18]. However, the problem is that they are an oxygen excess defect while annealing the film in an oxygen excess is known to reduce the instability [15,16]. A related model to the O vacancy model links the instability with the presence of In^{1+} valence sites whose In 6s lone pair states lie in the required energy range. However, this valence is not seen in the photoemission core level spectra [20].

A final possibility is that it is due to hydrogen. It has been proposed that single hydrogen atoms could undergo a deep to shallow transition and become donor-like [21,22]. However, this defect does not correspond to an oxygen deficiency and there is little other evidence for this mechanism. Thus, in this paper we propose a new mechanism in which the instability is

associated with a complex of two hydrogen atoms at an oxygen vacancy, a V_O/2H complex. This is an adaptation of the model of Du and Biswas [26] for hidden hydrogen in c-ZnO. We will show that this defect complex in the amorphous phase gives a consistent explanation of the bias stress instability and its persistent photo-conductivity. The complex gives rise to gap states lying just above the valence band maximum (VBM) as seen by photoemission, it clearly has an oxygen deficiency, it has the negative U property, and it has a low formation energy in the amorphous phase. In its neutral configuration, the hydrogens form 3-center bridging sites, whose vibrational signature has recently been seen experimentally in infra-red (IR) by Bang et al [27]. It should be noted that a considerable hydrogen content of over 10^{20} atoms cm^{-3} is seen in a-IGZO by thermal desorption spectroscopy or secondary ion mass spectrometry (SIMS) [28-30], which is many orders of magnitude more than that seen in crystalline (c-)ZnO. This concentration is similar to the density of sub-gap defect states seen in the amorphous films by photoemission [12], and much larger than the photo-excited carrier density created by $NBIS_8$.

Hydrogen has a complicated role in oxide semiconductors. In c-ZnO or a-IGZO, interstitial hydrogen acts as a shallow donor [31,32]. In this case the hydrogen is bonded to an oxygen atom as an OH bond. However, in the amorphous case, all the 10^{20} cm^{-3} hydrogens cannot be acting as donors, most must be compensated in some way. One way would be if the hydrogens were compensated by interstitial O acceptors. However, this would not have the necessary O deficiency. We show that the lowest energy of H pairs in a-ZnO is the V_O/2H state which is self-compensated for E_F above 2 eV. In this state, the hydrogens exist as anionic H^-, forming bridge bonds to metal atoms.

To show this, we study the incorporation of H_2 and water into an a-ZnO network. We use a-ZnO as a simplifying model of the a-IGZO network [32], as it has a higher symmetry crystalline phase, c-ZnO. We first create a random network of a-ZnO by density functional molecular dynamics using the CASTEP code [33,34], and then substitute an H_2 molecule for an oxygen site and study its different charge states. We then introduce an H_2O molecule into the *defect-free* network and study its behaviour. Finally, we consider the migration of the hydrogens from the dissociated water molecule to the oxygen vacancy. These results show that hydrogen or water can passivate the vacancy, but only at the expense of introducing states which lead to the possibility of the NBIS effect.

2. Methods

The calculations are carried out using 96 atom supercells of amorphous ZnO. These are created by a molecular dynamics (MD) anneal at 2000K for 20 ps, followed by a quench to 300K in 20 ps. The calculations are carried out using the CASTEP plane wave pseudopotential code. It uses norm-conserving pseudopotentials with a cutoff energy of 760 eV. The energies are converged to 10^{-5} eV per atom, and forces to under 0.01 eV/Å. Density functional theory (DFT) greatly underestimates the band gap of ZnO which can lead to errors, particularly in the description of the donor states of hydrogen. For static calculations, these can be rectified by using hybrid functionals such as the screened exchange functional for the electronic exchange-correlation functional. However, a less expensive method is desirable for the MD stage. For this, we use the GGA+U method. This includes an on-site potential U of 5 eV on the Zn 3d states to lower their binding energy [24,35]. However, this is not enough, and we also add a U potential of 5 eV on the O 2p states to further lower the energy of the valence band maximum [36]. The transition states of intrinsic defects are calculated using the Lany and Zunger [37] scheme.

The formation energy of each charge state is given by:

$$H_q(E_F,\mu) = [E_q - E_H] + q(E_v + \Delta E_F) + \sum_\alpha n_\alpha (\mu_\alpha + \Delta\mu_\alpha) \qquad (7.1)$$

where q is the charge on the system, E_q is the energy of charged system with a defect, E_H is the energy of charged defect-free system. E_V is the valence band maximum (VBM) and ΔE_F is the Fermi level with the respect to VBM. n_α is the number of atoms of species α, μ_α is the relative chemical potential of element α. We note that the first two terms are equal to the difference between the total energy of charged defect system and total energy of neutral defect-free system.

The density of c-ZnO is 5.44 g/cm^3, the density of a-ZnO is 4.62 g/cm^3[38]. The supercell of a-ZnO contains 96 atoms, and has a lattice constant of 11.3 Å. The O vacancy is created by removing one O atom from the supercell, and taking it externally to form half an O_2 molecule. The network or supercell is then relaxed by GGA. For a-ZnO, several oxygen atom sites are sampled to obtain an average defect formation energy. We take the average of the 6 lowest formation energies from 20 sampled sites (high formation energy sites will have a low probability.)

The water molecule is inserted into the bulk c-ZnO or a-ZnO network at a Zn-O bond. The structure is relaxed at constant volume.

3. Results and Discussion

Fig 7.1 shows the resulting 96 atom random network of ZnO. Most Zn and O atoms are 4-fold coordinated but some are 3-fold coordinated, reminiscent of the undercoordinated defects in GaAs [39,40].

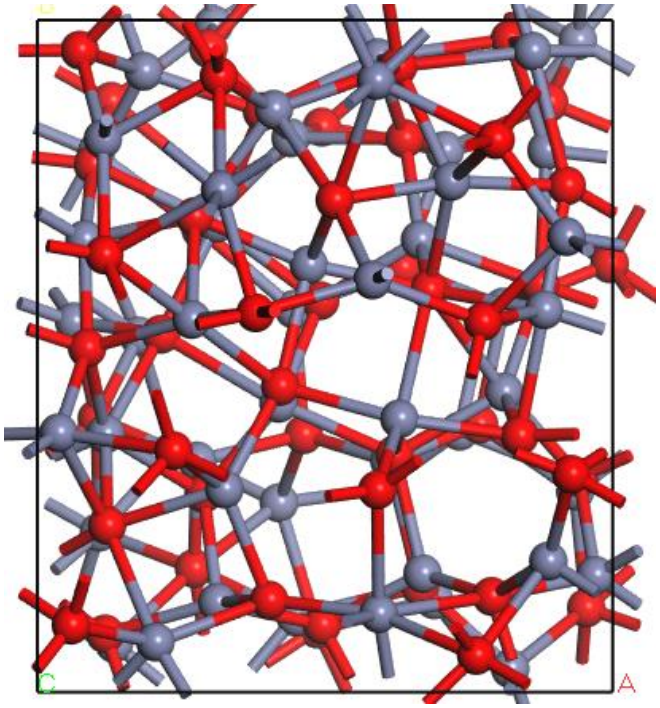


Fig 7.1 A random network of amorphous ZnO. Red balls = O, grey balls = Zn

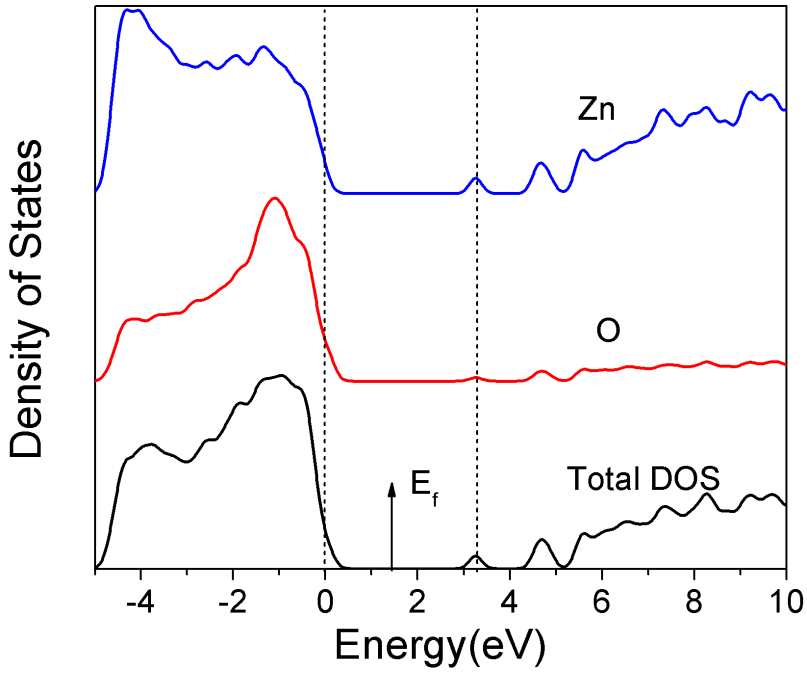
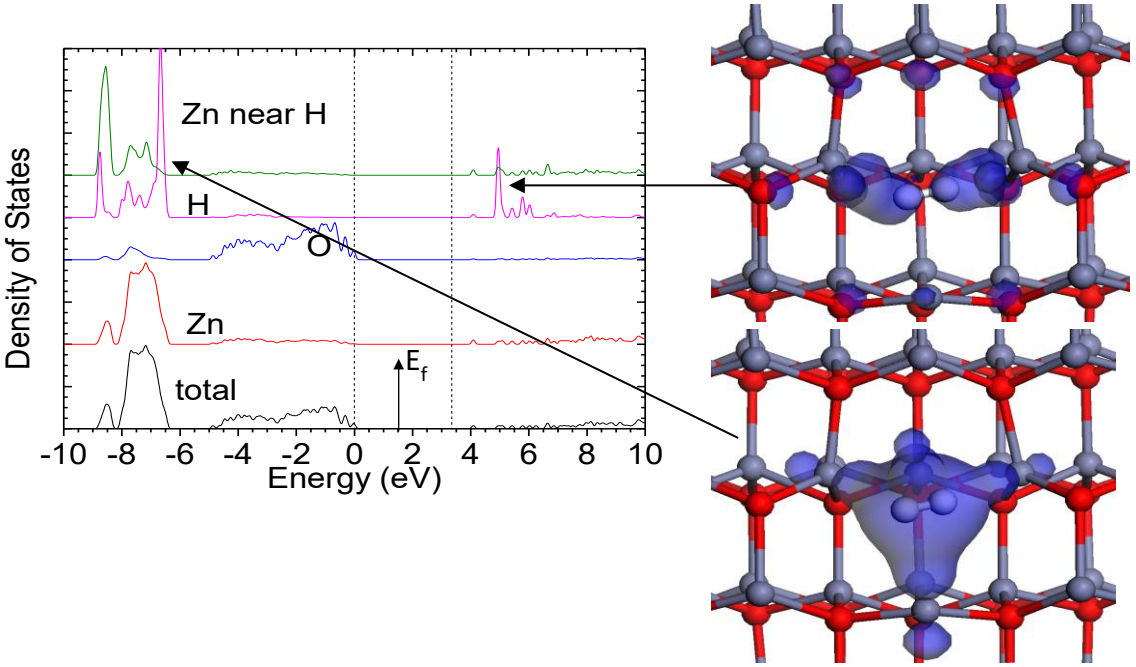


Fig 7.2 Partial density of states of the amorphous ZnO network in **Fig** 7.1

Fig 7.2 shows the density of states (DOS) calculated using the screened exchange (sX) hybrid functional [34]. sX corrects the band gap error of simple density functional theory. We see that the band gap is 3.2 eV, slightly smaller than that of the crystal.

3.1 Hydrogen in c-ZnO

We first consider the case of inserting H_2 into c-ZnO. **Fig 7.3(a,b)** shows an H_2 molecule at an O vacancy in ZnO. Its simplest configuration is the V_{O}^{2+}(2H) state. Here, the four metal dangling bond orbitals pointing into the vacancy are empty. This leaves two electrons on the two hydrogen atoms. These form a neutral H_2 molecule trapped inside the vacancy. This molecule gives rise to a filled HH bonding state lying at -6.5 eV below the ZnO valence band, whose wavefunction is shown in **Fig 7.3(b)**, and an empty anti-bonding state lying near +5.5 eV in its conduction band, as is seen in the partial density of states plots of **Fig 7.3(a)**. We find that a H_2 molecule at a vacancy has essentially the same total energy as it has in free space [26].

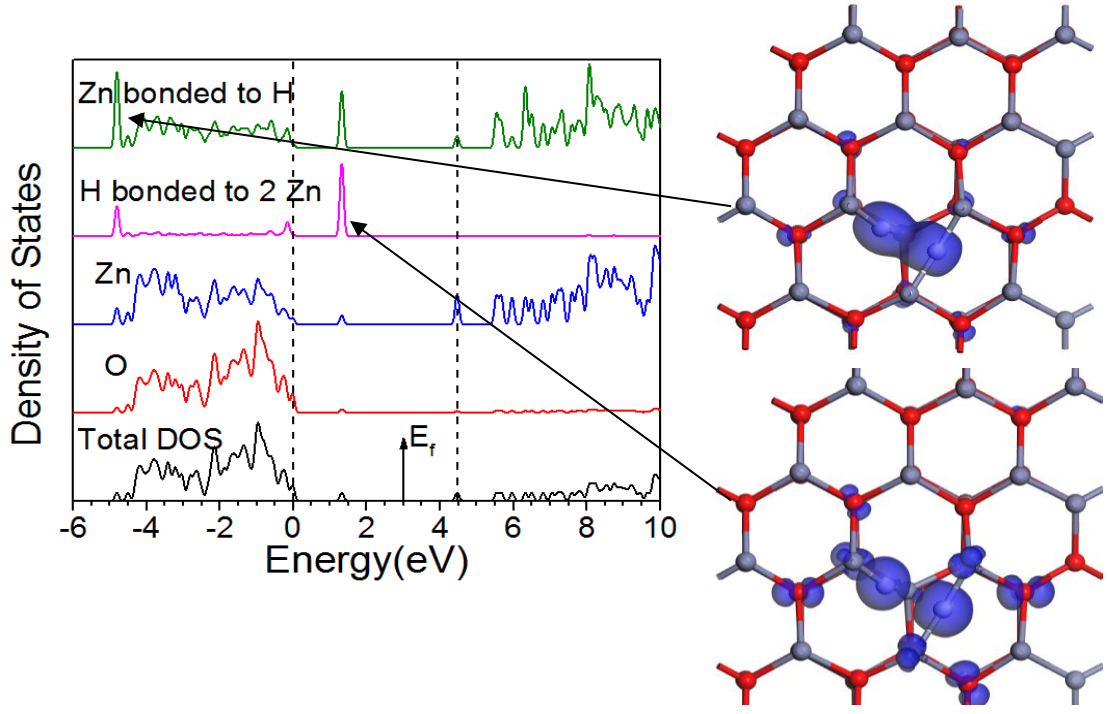


(a) (b)

Fig 7.3 (a) Partial density of states on the H site and on the adjacent Zn site, of the Vo^{2+} (2H) complex in crystalline ZnO with an internal H_2 molecule. (b) orbitals of the H-H bonding state at -6.5 eV (bottom) and the H-H antibonding state at +5.5eV (top). Red balls = O, grey balls = Zn, White balls = H.

The second configuration of interest is the neutral V_{O}^{0}(2H) state, as seen in **Fig 7.4(a,b)**. In this charge state, the complex has a completely different configuration of C_{2v} symmetry consisting

of two 3-center Zn-H-Zn bridges. Each bridge uses two of the four Zn dangling bond orbitals at the vacancy. Each bridge creates three states; a filled 3-center bonding state lying at –5.5 eV whose wavefunction is shown in **Fig 7.4**(b, upper panel), a filled 3-center state lying at +1.2 eV above the VBM, which is anti-bonding along the Zn-H bonds, whose wavefunction is shown in **Fig 7.4**(b, lower panel), and finally an empty Zn-Zn antibonding state lying well into the conduction band at roughly 5 eV. The energies of the states agree with those found by Du and Biswas[26]. We also note that the PDOS of the neutral V_O/2H complex in a-ZnO has a very similar energy spectrum to that in c-ZnO, comparing **Fig 7.4(a)** and **Fig 7.5**, showing that the bonding is similar in the two phases, although the gap state does lie slightly closer towards the valence band in a-ZnO.



(a) (b)

Fig 7.4. (a) Partial density of states of the neutral Vo^0 (2H) complex in c-ZnO, (b) orbitals of the bonding state at -5.5eV (top) and the gap state at 1.0 eV (bottom). This latter state is the cause of NBIS. The antibonding state around +5.5eV is unclear in PDOS. Red balls = O, grey balls = Zn, White balls = H.

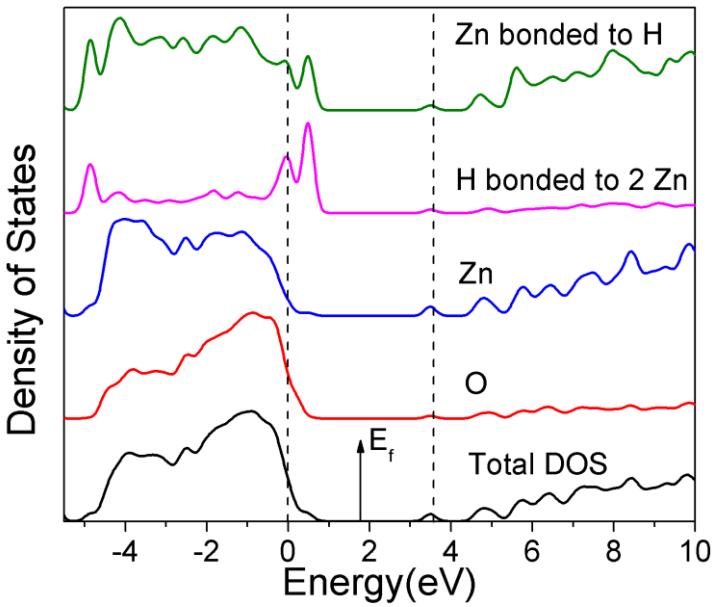


Fig 7.5 PDOS of the Vo^{0} (2H) complex in amorphous ZnO, equivalent to Fig 7.4(a).

A molecular orbital diagram showing the layout of these states is shown in **Fig 7.6**. The ideal O vacancy has T_d symmetry, and its vacancy orbitals form one A_1 and three T_2 states. When it acquires C_{2v} symmetry the T_2 states split into one of A_2 symmetry and a pair of B_2 symmetry. Meanwhile, the hydrogen orbitals form a symmetric and antisymmetric pair. The symmetric state interacts with the A_1 vacancy state and the antisymmetric one interacts with the A_2 vacancy state, each forming bonding and antibonding configurations. The four available electrons then fill the resulting two bonding states. The A_1 state lies at -5.5 eV and the A_2 state, being the key gap state, lies at +1.2 eV. The fact that both of these states are filled means that the H orbitals are overall 100% filled, giving the hydrogens an anionic H^- configuration [24]. As the hydrogens have an H configuration, the H^- pair is essentially iso-electronic with the O^{2-} ion.

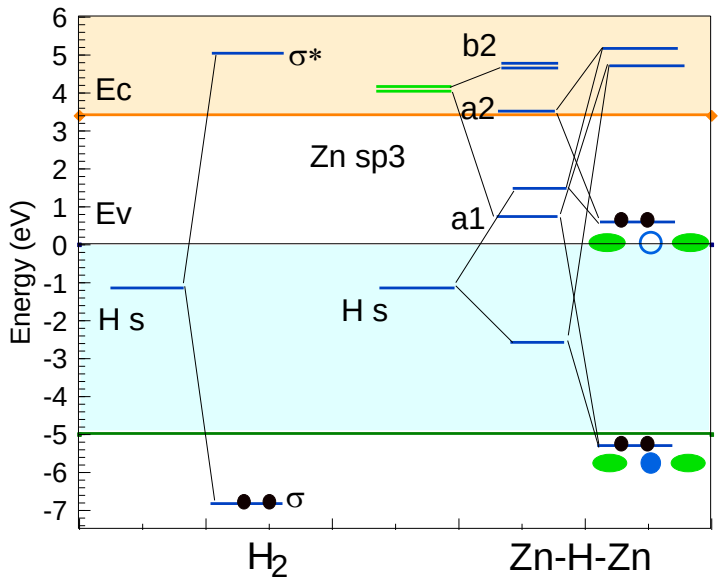


Fig 7.6. Approximate molecular orbital (MO) diagram of the Vo/2H defect in ZnO, in the 2+ state (left) and neutral state (right), showing the origin of the local orbitals.

Fig 7.7 shows the defect formation energies for the hydrogen interstitial donor and the Vo/2H complex in c-ZnO. As noted by Du and Biswas [26], the donor is the more stable configuration over the important range of Fermi energies (above midgap).

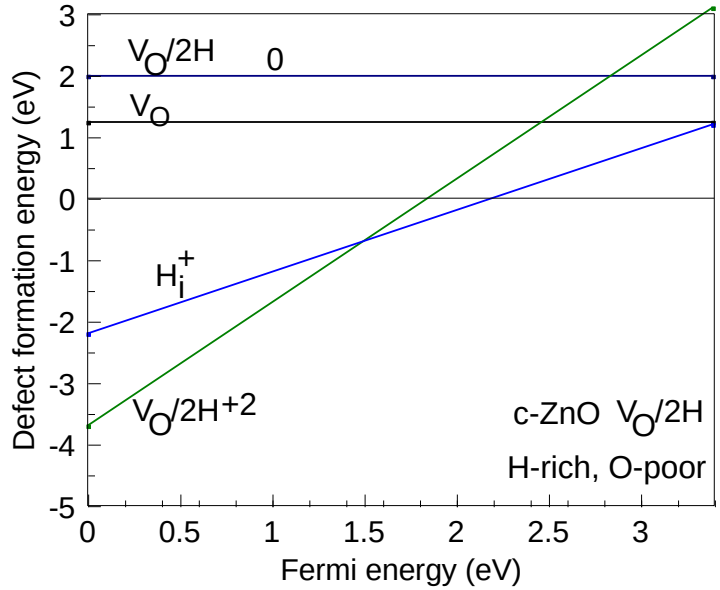


Fig 7.7 Defect formation energy vs. Fermi energy for the Vo/2H complex for the O-poor, H-rich condition in crystalling ZnO.

3.2 Hydrogen in a-ZnO

The average defect formation energy of the neutral O vacancy is +3.0 eV for a-ZnO in the O-rich condition (μ(O) = 0 eV), compared to +4.80 eV for c-ZnO, a sizable change. This is due to the lower density of a-ZnO which allows the atoms around the vacancy to relax more. +3.0 eV for O-rich corresponds to -0.34 eV for the O-poor condition.

Fig 7.8 shows the calculated defect formation energies of the VO/2H complex in a-ZnO for each charge state. We see that the 0 and +2 states are the most stable over the critical range of Fermi energies, above the midgap. The H^+ interstitial is less stable than the 2H/VO complex, over the range of Fermi energies, as seen in **Fig 7.8**. Thus, the complex is a negative U defect. This is different to the situation in c-ZnO. This explains the compensation of hydrogen in a-ZnO, and is significantly different to c-ZnO.

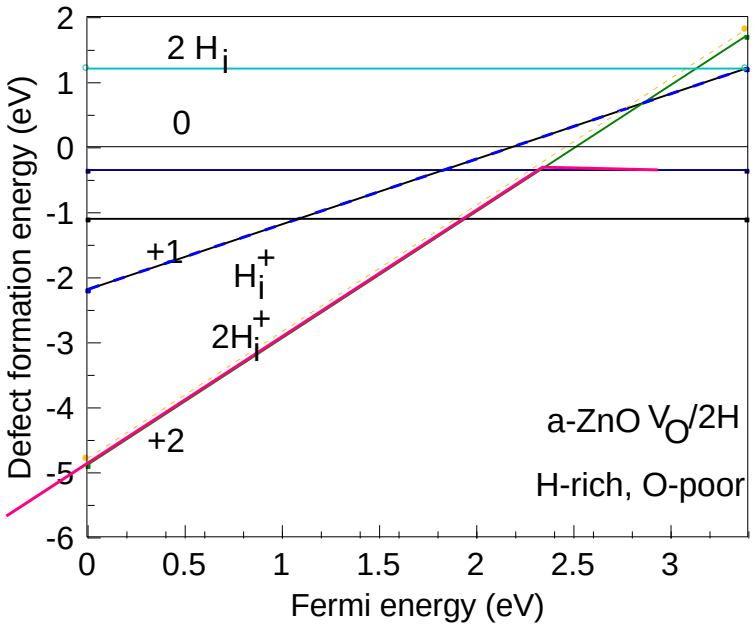
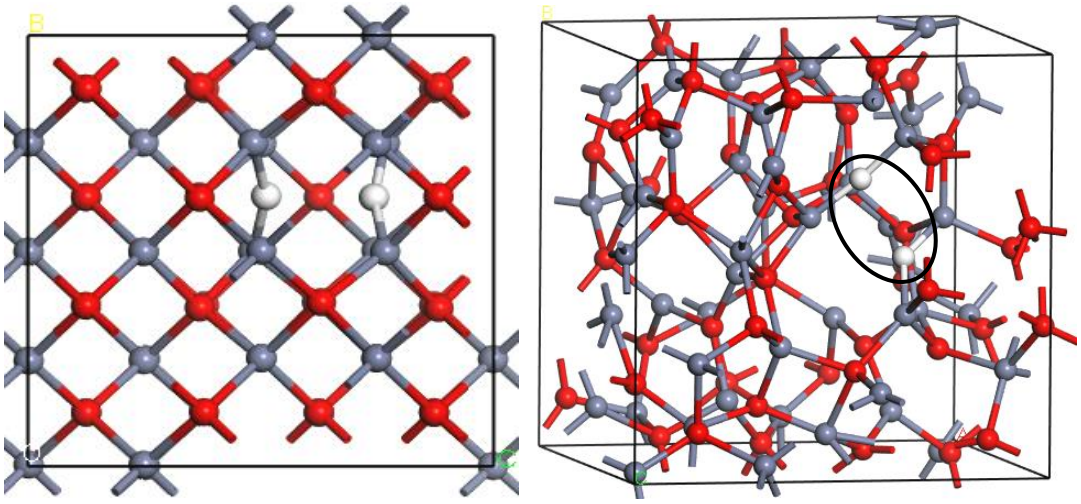


Fig 7.8 Defect formation energy vs. Fermi energy for Vo/2H complex and interstitial hydrogen in amorphous ZnO for O-poor condition.

The large positive defect formation energy of neutral VO/2H in c-ZnO means that the concentration of the VO/2H complex or 'hidden hydrogen' in c-ZnO is quite low, ~10^{16} cm^{-3}. The much lower formation energy of this defect in a-ZnO means that there is orders of magnitude higher H concentration in the amorphous phase, even for a lower formation temperature. In addition, the formation energy of the VO/2H complex in the H-rich condition

in a-ZnO is lower than that for the interstitial H donor configuration H_i in **Fig 7.8**, so that the compensated V_O/2H complex configuration is dominant.

Nomura et al [28] noted that IGZO typically contains a large concentration of hydrogen due to the poor vacuum pumping in the typical sputtering systems used for oxide deposition. Bang et al [27] identified the infra-red vibrational modes of the Zn-H-metal bridge groups in IGZO, and that they were more thermally stable than the vibrational modes of the O-H group. The greater stability of the Zn-H-metal bridge groups is because these defects are trapped at an O vacancy [41]. In contrast, the hydrogens at O-H bonds are interstitial, and migrate easily, as noted by Wardle et al [42].



(a) (b)

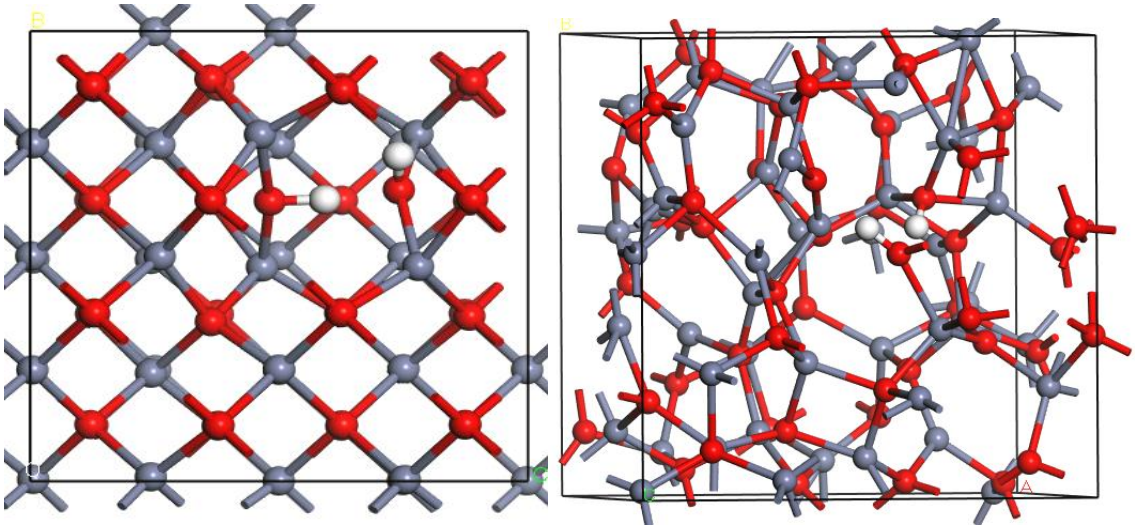
Fig 7.9. (a) The configuration of the Zn-H-Zn bridges from the V_O^0/2H complex in crystalline ZnO, (b) the equivalent configuration of (a) in amorphous ZnO, showing the conservation of local bonding (2H units circled). Red balls = O, grey balls = Zn, white balls = H.

Our calculations show that the states of metal-H bridge bonds are the typical cause of filled gap states at 1.0 eV above the valence band edge seen by X-ray photoemission [12]. Electrons in these states can be excited by photons of energy 2.5 eV into the conduction band and give the observed persistent photoconductivity [6,8,9]. These electrons can drift to the gate dielectric interface[7,9], giving a charge that causes a negative threshold voltage shift. It should be noted that the V_O^0(2H) / V_O^{2+}(2H) group also possesses a negative U, that is its 1+ state is less stable that the neutral or 2+ states. This energy barrier acts, as in the simple O vacancy [23,24], as a recombination barrier, so maintaining the persistent photoconductivity.

The $V_{O0}(2H)/V_{O2+}(2H)$ model has many similarities to the simpler O vacancy model, except that the origin of the photo-excited states is different. In the vacancy model, vacancy states are excited, which lies near midgap, whereas in the new $V_O^0(2H)/V_O^{2+}(2H)$ case, hydrogen bridge states are excited. These lie just above the valence band maximum and more consistent with photoexcitation by 2.5 eV photons. Thus, the present model is closer to the experimental situation.

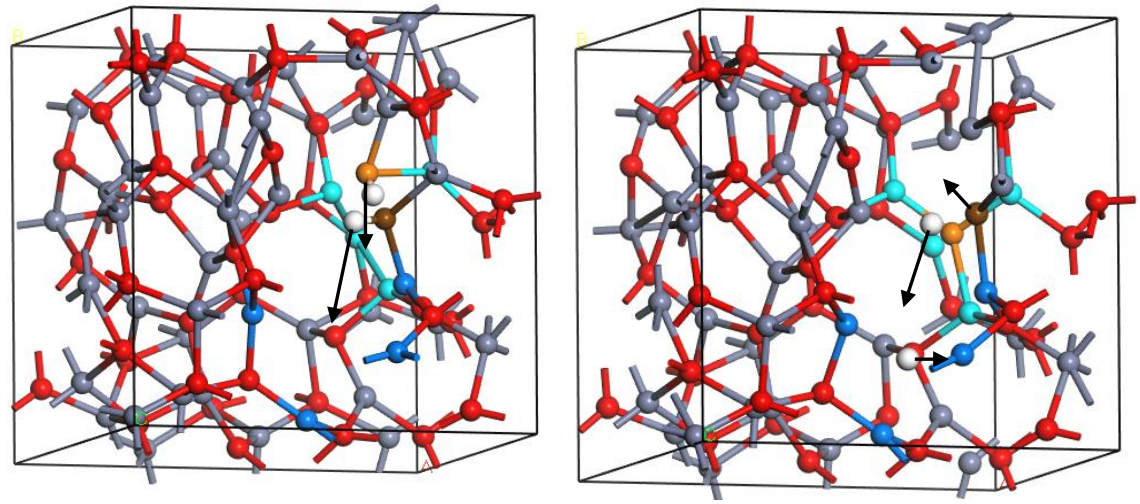
Overall, the details of the model have been calculated for the binary oxide a-ZnO, but a similar situation applies to other oxides like a-IGZO, $ZnSnO_x$ and In_2O_3 etc, which all have similar band gaps. ZnO has a 4-fold covalent random network, while the other oxides have more ionic bonding associated with the In or Sn sites. Nevertheless the behaviour of H at an O vacancy is similar in each case, **Fig 7.9(a,b)**, in that it gives rise to Zn-H-Zn bridges with filled states in the lower band gap region, which can be photo-excited.

3.3 H_2O in c-ZnO and a-ZnO

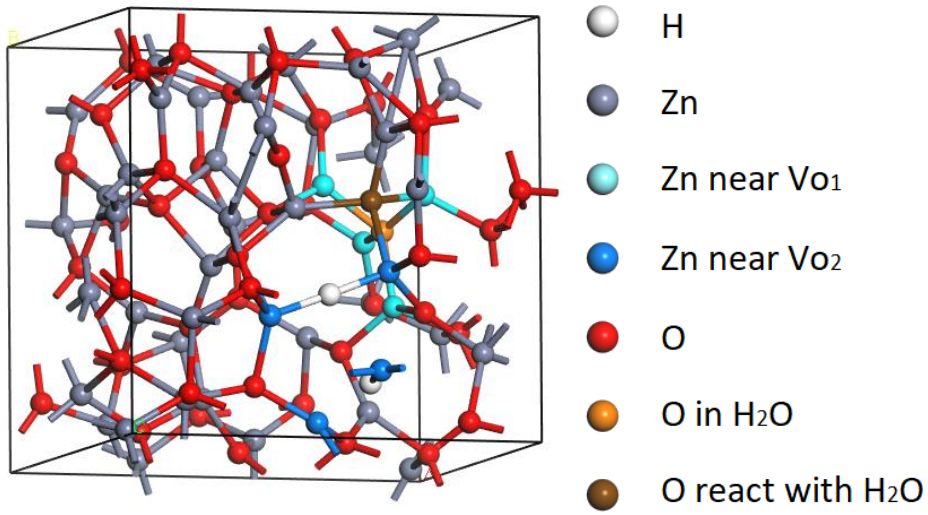


(a) (b)

Fig 7.10 (a) The configuration of dissociated water molecule in crystalline ZnO, (b) the equivalent configuration of (a) in amorphous ZnO. The hydrogen atoms are bonded covalently to one oxygen and weakly to another oxygen. Red balls = O, grey balls = Zn, white balls = H.



(a) (b)



(c)

Fig 7.11 (a) A dissociated water molecule in amorphous ZnO, adjacent to two oxygen vacancies, (b) the transition state where two hydrogens are moving to one of the vacancies (moving direction shown with arrows), and the oxygen is moving to the other vacancy, (c) final state in which the water's oxygen has reached one of the vacancies, and the two hydrogen atoms have reached the other oxygen vacancy, thereby passivating the two vacancies. Blue and turquoise balls are the Zn atoms adjacent to each of the O vacancies for guidance, white balls = H, as noted in the color legend.

We finally consider the addition of a water molecule to the ZnO network. **Fig 7.10(a,b)** shows the water in crystalline and amorphous ZnO. The H_2O species dissociates into H and OH groups which insert into a Zn-O bond. The OH group attaches itself to the metal dangling bond

and the H attaches to the O dangling bond, resulting in two adjacent –OH groups. The OH groups remain close to each other, bound by a weak hydrogen bonding. In effect, each H is bonded strongly to one O site and weakly bonded to the second O site in an asymmetric bridge configuration resembling that the hydrogen sites in water.

One H_2O group in ZnO consists of an excess of a H_2 and an O compared to the pure state. Therefore, it is intrinsically possible for one H_2O molecule to passivate two O vacancy. **Fig 7.11(a)** shows the initial state of such process with the $(ZnO)_n(H_2O)_1$ configuration plus two nearby neutral O vacancies. **Fig 7.11(b,c)** shows how two hydrogens from the H_2O group could move to one of the vacancies and the O of the H_2O could move to the other vacancy, via the transition state of **Fig 7.11(b)**. In the final state of **Fig 7.11(c)**, the O has filled one of the O vacancies, and thus passivated it. The two hydrogens have filled the other vacancy, and bonded to all the Zn dangling bonds. They have passivated it in the sense that they have removed the deep mid-gap state associated with a neutral O vacancy in ZnO. They have replaced its mid-gap state with states close to the VBM associated with the Zn-H-Zn bridges. The process has removed states in the upper half of the gap and mid-gap traps that affect the operation of the oxide as an n-type transistor. Thus, the water molecule has passivated states affecting the transistor action, as seen experimentally [41,42].

3.4 Comparison of hydrogen behaviours in a-ZnO and a-Si:H

Finally it is interesting to compare the behaviour of hydrogen in a-ZnO with that a-Si:H. Although hydrogen passivates the single Si dangling bond in a-Si:H [43], it seems that hydrogen occurs mostly in pairs in Si or a-Si:H [44-46]. For example, in crystalline Si, H will form the H_2^* defect, which is a low energy hydrogen configuration caused by inserting two hydrogens into a Si-Si bond [41]. This two-hydrogen defect has an even lower energy in the a-Si:H network, and is also strongly involved in the network growth process of a-Si:H [46]. It is the main defect in the instability mechanisms of a-Si:H [47-48]. In the same way, the two-hydrogen defect is more stable in a-ZnO than in c-ZnO and this causes its leading role in the amorphous oxide semiconductors [49-50].

4. Summary

In summary, it is proposed that the sizable hydrogen content seen in amorphous oxide semiconductors exist as hydrogen atom pairs trapped at oxygen vacancies. These form metal-H-metal bridge bonds in oxide semiconductors and are the source of the negative bias

illumination stress instability observed in these materials. Our calculations show that such bonds give rise to defect states in the lower gap region which are the cause of this instability.

5. References

[1] Nomura, K, *et al*, H, Room-temperature fabrication of transparent flexible thin-film transistors using amorphous oxide semiconductors. *Nature* **432** 488-492 (2004)

[2] Hosono, H., Ionic amorphous oxide semiconductors, material design, carrier transport and device application. *J. Non. Cryst. Solids* **352,** 851 (2006)

[3] Kamiya, T. & Hosono, H., Material characteristics and applications of transparent amorphous oxide semiconductors. *NPG Asia Mater* **2,** 15 (2010)

[4] Fortunato, E., Barquinha, P., Martins, R., Oxide Semiconductor Thin-Film Transistors. *Adv Mater* **24** 2945-2986 (2010);

[5] Park, J. S., Maeng, W.S., Kim, H.S., & Park, J.S., Review of recent developments in amorphous oxide semiconductor thin film transistor devices. *Thin Solid Films* **520,** 1679 (2012)

[6] Goerrn, P., Lehnhardt, M., Riedl, T., Kowalsky, W., The influence of visible light on transparent zinc oxide thin film transistors. *Appl. Phys. Lett*. **9**1, 193504 (2009)

[7] Ghaffarzadeh, K., *et al*, Instability in the threshold voltage and subthreshold behaviour in Hf-In-Zn-O thin film transistors induced by bias and light stress. *Appl. Phys. Lett.* **97**, 113504 (2010), Persistent Photoconductivity in Hf-In-Zn-O thin film transistors. *Appl. Phys. Lett.* **9**7, 143510 (2010)

[8] Nomura, K., Kamiya, T., Hosono, H., Interface and bulk effects for bias-light-illumination instability in amorphous In-Ga-Zn-O thin film transistors. *J. Soc. Inf. Disp.* **18,** 789 (2010)

[9] Jeong, J. K., Photo-bias instability of metal oxide thin film transistors for advanced active matrix displays. *J Mater. Res.* **28**, 2071 (2013)

[10] Chowdhury, M.D.H., Migliorato, P. & Jang, J., Light-induced instabilities in amorphous In-Ga-Znoxide thin film transistors. *Appl. Phys. Lett*.**97,** 173506 (2010)

[11] Jeon, S., et al, Gated three-terminal device architecture to eliminate persistent photoconductivity in oxide semiconductor photosensor arrays. *Nature Mater*. 11 301 (2012)

[12] Nomura, K., *et al*, Subgap states in transparent amorphous oxide semiconductor In-Ga-Zn-O observed by bulk sensitive x-ray photoelectron spectroscopy. *Appl. Phys. Lett.* **92,** 202117 (2008)

[13] Ryu, B., *et al*., O-vacancy as the origin of negative bias illumination stress instability in amorphous InGa-Zn-O thin film transistors. *Appl. Phys. Lett* **97,** 022108 (2010);

[14] No, H.K., *et al*, Electronic structure of oxygen-vacancy defects in amorphous In-Ga-Zn-O semiconductors. *Phys Rev B* **84,** 1115205 (2011)
[15] Ji, K. H., *et al*, Effect of high pressure oxygen annealing on negative bias illumination stress induced instability in InGaZnO thin film transistors. *Appl Phys. Lett*. **98,** 103509 (2011)
[16] Yang, S., *et al*, Suppression of negative bias illumination instability of Zn-Sn-O transistor using oxygen plasma treatment. *Appl. Phys. Lett.* **99,** 102103 (2011)
[17] Nahm, H.H., Kim, Y.S. & Kim, D.H., Instability of amorphous oxide semiconductors via carriermediated structural transition between disorder and peroxide state. *Phys Stat Solidi B* **249,** 1277 (2012)
[18] Robertson, J. & Guo, Y., Light induced instability mechanism in amorphous In-Ga-Zn-O semiconductors. *Appl. Phys. Lett* .**104,** 162102 (2014)
[19] Walsh, A., DaSilva, J.L.F., & Wei, S. H., Interplay between order and disorder in the high performance of amorphous transparent conducting oxides. *Chem Mater* **21,** 5119 (2009)
[20] Sallis, S., *et al*., Deep subgap feature in amorphous indium gallium zinc oxide:evidence against reduced indium. *Phys. Stat. Solidi.A* **212,** 1461 (2015)
[21] Nahm, H. H., Park, C.H., & Kim, S. H., Bistability of hydrogen in ZnO, origin of doping limit and persistent photoconductivity. *Sci. Reports* **4** 4124 (2014)
[22] Kang, Y., *et al*., Hydrogen bistability as the origin of photo-bias-thermal instabilities in amorphous oxide semiconductors. *Adv. Electronic Mater.* **1,** 1400006 (2015)
[23] Lany S. & Zunger, A., Anion vacancies as a source of persistent photoconductivity in II-VI and chalcopyrite semiconductors. *Phys. Rev. B* **72,** 035215 (2005)
[24] Janotti, A. & Van de Walle, C.G., Oxygen vacancies in ZnO. *Appl. Phys. Lett.* **87,** 122102 (2005)
[25] Oba, F., Togo, A., Tanaka, I., Paier, J., Kresse, G., Defect energetics in ZnO: hybrid Hartree-Fock density functional study. *Phys. Rev. B* **77,** 245202 (2008)
[26] Du, M.H., & Biswas, K., Anionic and hidden hydrogen in ZnO. *Phys Rev Lett* **106,** 115502 (2011)
[27] Bang, J., Matsuishi, S., Hosono, H., Hydrogen anion and subgap states in amorphous In-Ga-Zn-O thin films for TFT applications. *Appl Phys Lett* **110,** 232105 (2017)
[28] Nomura, K., Kamiya, T., & Hosono, H., Effects of diffusion of hydrogen and oxygen on electrical properties of amorphous oxide semiconductor, In-Ga-Zn-O. *ECS J Solid State Sci. Technol*. **2,** 5 (2013)

[29] Miyase, T., *et al*., Roles of hydrogen in amorphous oxide semiconductor in In-Ga-Zn-O, comparison of conventional and ultra-high vacuum sputtering. *ECS Solid State Sci. Technol.* **3,** Q3085 (2014)

[30] Hanyu, Y., et al, Hydrogen passivation of electron trap on amorphous In-Ga-Zn-O thin film transistor. *App Phys Lett* **103,** 202114 (2013)

[31] Van de Walle, C. G., Hydrogen as a cause of doping in ZnO. *Phys. Rev. Lett* . **85,** 1012 (2000)

[32] Omura, H., *et al*, First principles study of defects in crystalline indium gallium zinc oxide. *J. App. Phys.* **105**, 093712 (2009)

[33] Clark, S.J., *et al*., First principles methods using CASTEP. *Z. Kristallogr* **220**, 567 (2005)

[34] Clark, S. J., & Robertson, J., Screened exchange density functional applied to solids. *Phys. Rev. B* **82** 085208 (2010)

[35] Lany, S., & Zunger, A., Dopability, intrinsic conductivity and non-stoichiometry of transparent conducting oxides. *Phys. Rev. Lett.* **98,** 045501 (2007)

[36] Lany, S., & Zunger, A., Polaronic hole localization and multiple hole binding of acceptors in oxide widegap semiconductors. *Phys. Rev. B* **80,** 085202 (2009)

[37] Lany, S., & Zunger, A., Assessment of correction methods for the band-gap problem and finite-size effects in supercell defect calculations: Case studies for ZnO and GaAs *Phys. Rev. B* **78**, 235104 (2008)

[38] Kingey W. D., Mickelson R. A., and Bellevue W., "Amorphous zinc oxide semiconductor and method of making," U.S. patent (December 27, 1966)

[39] O'Reilly E.P., & Robertson, J., Electronic structure of amorphous III-V and II-VI compound semiconductors and their defects. *Phys. Rev. B* **34,** 8684 (1986)

[40] Fois, E., et al, Structure, electronic properties and defects of amorphous GaAs. *Phys. Rev. B* **45,** 13378 (1992)

[41] Janotti, A., & van de Walle, C.G., Hydrogen multicenter bonds. *Nature Mats* **6,** 44 (2007)

[42] Wardle, M.G., Goss, J.P., & Briddon, P., First principles study of diffusion of hydrogen in ZnO. *Phys. Rev. Lett.* **96,** 205504 (2006)

[43] Nomura, K., *et al*., Defect passivation and inhomogenization of amorphous oxide thin-film transistor by wet O_2 annealing. *App. Phys. Lett.* **93** 192107 (2008)

[44] Nayak, P.K., et al, Six-fold improvement of In-Zn-O thin film transistors using a simple water treatment. *Adv. Elec. Mats.* **1,** 1500014 (2015)

[45] Paul, W., *et al*, Doping, Schottky barrier and p-n junction formation in a-Ge and Si by sputtering. *Solid State Commun.* **20** 969 (1976)

[46] Chang, K.J., & Chadi, D.J., Hydrogen bonding and diffusion in crystalline silicon. *Phys. Rev. B* **40**, 11644 (1989)

[47] van de Walle, C.G., Energies of various configurations of hydrogen in silicon. *Phys. Rev.* B **49,** 4579 (1994)

[48] Robertson, J., Growth mechanism of hydrogenated amorphous silicon. *J. App. Phys.* **87,** 2608 (2000)

[49] Biswas, R., Li, Q., Pan, C., & Yoon, Y., Mechanism for hydrogen diffusion in amorphous silicon. *Phys. Rev. B* **57,** 2253 (1998)

[50] Deane, S.C., & Powell, M.J., Defect-pool model and hydrogen density of states in hydrogen amorphous silicon. *Phys. Rev. B* **53,** 10121 (1996)

Chapter 8. Conclusion and Outlook

In summary, this work focuses on simulating defect behaviours in oxides for applications in MOSFET and ReRAM. Oxide play an important role in these devices, and the performance of these devices are determined by the defect behaviours in the oxide layers. Oxide act as the insulating layer in MOSFET, which has an essential impact on gate electric control and carriers' behaviours in channel. Defects, especially oxygen vacancy, in oxide and oxide/channel interface not only cause charge trapping which impedes gate voltage swing, but also results in device reliability problem. On the other hand, aggregation of defects in oxide creates conductive filament in ReRAM, but how this process happen in atomic scale is difficult to investigate in experiment. Density functional theory simulation provides a powerful tool to investigate the underlying physics for those existing problems, and shines the light on possible solution. Hybrid functional is useful to accurately simulate the electronic states of defects and band alignment between different materials in real devices without introducing any spurious electronic interaction. We have applied these tools to study the defect behaviours in functional oxides of the related electronic devices. Not only formation process of defects is covered, but also defect diffusion process and relative stability are included in this research, which is of great importance to the reliability of MOSFET. Moreover, doping of oxide is necessary and unavoidable in the electronic industry. We have applied the concept of defect compensation/passivation in oxides by keeping the closed sell electronic configuration of host material, which is an application of electronic counting rule. Successful control of doping in oxides can either passivate existing unwanted defects or manipulate defect behaviours according to demands. These problems require understanding of underlying physical and chemical process in devices, and this study tries to make certain contribution to defect related problems in oxides.

Interface is highly involved in each aspect of this topic. From device level, interface is always at the core of working mechanism. For MOSFET, the conducting channel is near interface region and greatly affected by interface quality. Interfacial traps and border traps can impede channel carriers which directly determines the performance of the whole device. For ReRAM, the metal/oxide interface is the pool of oxygen defects, which contributes to both set and reset working mechanism. From material level, the band alignment or band offset between oxide and semiconductor is one of the fundamental selection rules for oxides in MOSFET. In addition, the Schottky barrier height for the interface between metal and semiconductor is difficult to tune due to the interfacial defect states, which is covered in chapter 5. Moreover, growth of

native oxide on top of semiconductor layer start from interface region, where chemical environment makes the mechanism totally different from that in bulk oxide, as investigated in chapter 4. Furthermore, interface is the region where crystal loses the high symmetry in structure. Therefore, defects usually have a much higher density in this region, and any problems like doping can have totally different influence at interface compared to other region. Finally, Ge MOSFETs are known to suffer from defect related interface quality problem, including high D_{it}, sub-oxide interfacial layers, and reliability. In this work, we have addressed these problems and put forward explanations from simulation view.

For future research, efforts could be further focused on oxide choice for Ge MOSFET. GeO_2 oxidation mechanism has been proved to be different from SiO_2, which intrinsically involves more defects. Doping in GeO_2 with rare earth element somehow improve the reliability, of which the mechanism has been explained in this study. However, the device performance is still not satisfactory. Other choice of oxide and doping scheme is worthwhile to investigate in search of the oxide most compatible with Ge technology. Moreover, it is still unclear why Ge tend to react with HfO_2 while Si works with HfO_2 well. Some molecular dynamic simulation is needed to understand the atom diffusion behaviours at these two interfaces.

Table of Acronym

AMLCD	Active Matrix Liquid Crystal Display
AOS	Amorphous Oxide Semiconductor
CASTEP	CAmbridge Serial Total Energy Package
CB	Conduction Band
CBM	Conduction Band Minimum
CMOS	Complementary Metal Oxide Semiconductor
CNL	Charge Neutrality Level
CV	Capacitance-Voltage
DB	Dangling Bond
DFT	Density Functional Theory
D_{it}	Interfacial Trap Density
DOS	Density of States
EOT	Equivalent Oxide Thickenss
ESR	Electron Spin Resonance
FET	Field Effect Transform
FLP	Fermi Level Pinning
GGA	Generalised Gradients Approximation
HSE	Heyd-Scuseria-Ernzerhof
HRS	High Resistance State
HRTEM	High Resolution Transmission Electron Microscopy
IGZO	In-Ga-Zn-O
LDA	Local Density Approximation
LRS	Low Resistance State
MD	Molecular Dynamics
MIGS	Metal Induced Gap State
MOSFET	Metal Oxide Semiconductor Field Effect Transistor
NBIS	Negative Bias Illumination Stress

OLED	Organic Light Emitting Diode
PBE	Perdew-Burke_Ernzerhof
PDOS	Partial Density of States
PPC	Persistent PhotoConductivity
ReRAM	Resistive Random Access Memory
RS	Resistance Switching
SBH	Schottky Barrier Height
sX	Screened Exchange
TFT	Thin Film Transistor
VAP	Valence Alternation Pair
VB	Valence Band
VBM	Valence Band Maximum
WF	Work Function
XPS	X-ray Photoemission Spectroscopy